COMPUTATIONALLY EFFICIENT APPROACHES FOR MODERN OPTIMIZATION PROBLEMS

SHIYI JIANG

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Computationally Efficient Approaches for Modern Optimization Problems

Shiyi Jiang

A thesis submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

September 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

Signature: ____________________

Name of Student: Shiyi Jiang

Abstract

Optimization problems lie at the core of decision-making processes across multiple disciplines, including supply chain management, industrial engineering, and energy systems. With the exponential growth of data volume and increasing complexity of industrial applications, these optimization problems have grown substantially in both scale (characterized by the number of decision variables and constraints) and complexity (reflected in non-convex, stochastic, or robust formulations). Conventional solution approaches often prove inadequate when addressing such challenges, frequently encountering computational bottlenecks due to either time or memory constraints. This growing computational challenge has created an urgent need for developing more efficient solution approaches capable of handling modern large-scale and complex optimization problems. In this thesis, we present three recent studies on computationally efficient approaches for modern optimization problems.

In this first study, we investigate moment-based distributionally robust optimization (DRO), where the underlying joint distribution of random parameters runs in a distributional ambiguity set constructed by moment information. Although most moment-based DRO problems can be reformulated as semidefinite programming (SDP) problems that can be solved in polynomial time, solving high-dimensional SDPs is still time-consuming. Unlike existing approximation approaches that first reduce the dimensionality of random parameters and then solve the approximated SDPs, we propose an optimized dimensionality reduction (ODR) approach by integrating the dimensionality reduction of

random parameters with the subsequent optimization problems. Such integration enables two outer and one inner approximations of the original problem, all of which are low-dimensional SDPs that can be solved efficiently, providing two lower bounds and one upper bound correspondingly. More importantly, these approximations can theoretically achieve the optimal value of the original high-dimensional SDPs. As these approximations are nonconvex SDPs, we develop modified Alternating Direction Method of Multipliers (ADMM) algorithms to solve them efficiently. We demonstrate the effectiveness of our proposed ODR approach and algorithm in solving multiproduct newsvendor and production-transportation problems. Numerical results show significant advantages of our approach regarding computational time and solution quality over the three best possible benchmark approaches. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude.

In the second study, we consider non-convex quadratically constrained programs (QCPs), which are generally NP-hard and challenging problems. We propose two novel mixed-integer linear programming (MILP) approximations for a non-convex QCP. Our method begins by utilizing an eigenvalue-based decomposition to express the non-convex quadratic function as the difference of two convex functions. We then introduce an additional variable to partition each non-convex constraint into a second-order cone (SOC) constraint and the complement of an SOC constraint. We employ two polyhedral approximation approaches to approximate the SOC constraint. The complement of an SOC constraint is approximated using a combination of linear and complementarity constraints. As a result, we approximate the non-convex QCP with two linear programs with complementarity constraints (LPCCs). More importantly, we prove that the optimal values of the LPCCs asymptotically converge to that of the original non-convex QCP. By proving the boundedness of the LPCCs, we further reformulate the LPCCs as MILPs. We demonstrate the effectiveness of our approaches via numerical experiments by applying our proposed approximations to randomly generated instances and two application problems: the joint decision and estimation problem and the two-trust-region subproblem. The numerical results show significant advantages of our approaches in terms of solution quality and computational time compared to existing benchmark approaches.

In the third study, we focus on the planning of distributed energy resources in distribution systems. This problem has been challenged by the significant uncertainties and complexities of distribution systems. To ensure system reliability, one often employs chance-constrained programs to seek a highly likely feasible solution while minimizing certain costs. The traditional sample average approximation (SAA) is commonly used to represent uncertainties and reformulate a chance-constrained program into a deterministic optimization problem. However, the SAA introduces additional binary variables to indicate whether a scenario sample is satisfied and thus brings great computational complexity to the already challenging distributed energy resource planning problems. In this study, we introduce a new paradigm, i.e., the partial sample average approximation (PSAA) using real data, to improve computational tractability. The innovation is that we sample only a part of the random parameters and introduce only continuous variables corresponding to the samples in the reformulation, which is a mixed-integer convex quadratic program. Our extensive experiments on the IEEE 33-Bus and 123-Bus systems show that the PSAA approach performs better than the SAA because the former provides better solutions in a shorter time in in-sample tests and provides better guaranteed probability for system reliability in out-of-sample tests.

Publications Arising from the Thesis

1. Shiyi Jiang, Jianqiang Cheng, Kai Pan, and Zuo-Jun Max Shen, "Optimized Dimensionality Reduction for Moment-based Distributionally Robust Optimization", in *Operations Research*.

2. Shiyi Jiang, Jianqiang Cheng, Kai Pan, Boshi Yang, "Asymptotically tight MILP approximations for a nonconvex QCP", in *INFORMS Journal on Computing*.

3. Shiyi Jiang, Jianqiang Cheng, Kai Pan, Feng Qiu, and Boshi Yang, "Data-Driven Chance-Constrained Planning for Distributed Generation: A Partial Sampling Approach", in *IEEE Transactions on Power Systems*.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr. Kai Pan, for his unwavering guidance, support, inspiration, and encouragement throughout my journey at PolyU. It has been both an honor and a privilege to learn from him. Beyond his exceptional scientific mentorship, he has been a remarkable role model in all aspects of life—for which I am endlessly grateful.

My sincere appreciation also goes to Dr. Jianqiang Cheng, Dr. Boshi Yang, Dr. Zuo-Jun Max Shen, and Dr. Feng Qiu, whose collaboration and invaluable insights have significantly contributed to my growth as a researcher. I am honored to have Dr. Miao Song, Dr. Huifu Xu, and Dr. Zhi Chen in my thesis committee. I am very grateful to them for their valuable comments and remarkable advice to improve this thesis.

Being part of the Department of Logistics and Maritime Studies at PolyU has been a tremendous privilege. I extend my thanks to the entire academic community for their support during my PhD studies. Special thanks go to Mr. Silong Zhang and Mr. Tao Zhang for their emotional encouragement—their presence made challenges easier to overcome. I am also grateful to my friends in offices MN037 and MN038. Their companionship turned my time in Hong Kong into an unforgettable experience.

Finally, no words can fully capture my gratitude to my family, especially my parents, whose unconditional love and patience have been my steadfast foundation throughout this journey.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Mathematical optimization serves as a fundamental tool for decision-making across numerous disciplines, from supply chain management to energy system operations. In contemporary industrial applications, two critical trends significantly increase the computational challenges in optimization problems: (i) the explosive growth in problem scale, with modern applications often involving millions of decision variables and constraints (Bertsimas et al. 2019a, Xavier et al. 2021), and (ii) the increasing complexity of problem formulations, particularly with the incorporation of non-convexity (Dong and Luo 2018, Belotti et al. 2009) and uncertainty (Birge and Louveaux 2011, Rahimian and Mehrotra 2022).

These challenges create an urgent need for more computationally efficient solution approaches. In this thesis, we develop such approaches for three challenging problems: (i) the high-dimensional semidefinite programming (SDP) formulation from moment-based distributionally robust optimization (DRO), (ii) the non-convex quadratically constrained program (QCP), and (iii) the chance-constrained formulation from the power grid planning problem.

The first challenging problem arises from moment-based DRO. DRO is a modeling framework that integrates statistical information with traditional optimization methods (Scarf 1958, Delage and Ye 2010). Under this framework, the underlying joint distribution of random parameters is assumed to belong to a distributional ambiguity set inferred from given data or prior belief. Decisions are then optimized against the worst-case distribution within the set (see Rahimian and Mehrotra 2019 and Lin et al. 2022 for detailed reviews). Moment-based DRO considers moment-based ambiguity sets, which include all distributions satisfying certain moment constraints. Examples of such constraints include restricting the exact mean and covariance matrix (Scarf 1958) and bounding the first and second moments (Ghaoui et al. 2003, Delage and Ye 2010). Although most moment-based DRO problems can be reformulated as SDP problems solvable in polynomial time using traditional interior-point methods, high-dimensional SDPs impose substantial computational time and memory demands (Vandenberghe and Boyd 1996, Helmberg 2002). Widely adopted commercial solvers (e.g., Mosek) exhibit prohibitively long computational times when solving high-dimensional SDPs, with the burden escalating rapidly as problem dimension increases. Thus, it is of great interest to study efficient algorithms for solving SDPs in the context of moment-based DRO (Chapter 2).

The second challenging problem considered in this thesis is the non-convex QCP, characterized by non-convex quadratic constraints. It has attracted substantial attention due to its extensive applications across industries, including signal processing (Huang and Palomar 2014, De Maio et al. 2010), portfolio optimization (Ban et al. 2018), wireless communications (Zhang and So 2011, Xiang and Tao 2012), and machine learning (Zien and Ong 2007, Bach et al. 2004). For example, the AC optimal power flow (ACOPF) problem in power system operations—a fundamental industrial problem—has been extensively studied and reformulated as a non-convex QCP (Cain et al. 2012). Similarly, the joint decision and estimation (JDE) problem in signal processing can also be reformulated as a non-convex QCP (Zang and Zhu 2022b). Solving non-convex QCPs remains a significant challenge due to their NP-hard nature. Various optimization solvers (Sahinidis 1996, Tawarmalani and Sahinidis 2004, Belotti et al. 2009, Lin and Schrage 2009, Misener and Floudas 2013, 2014, Bestuzheva et al. 2021) have been developed to tackle them. Among these, the Gurobi solver (Gurobi 2024) is currently one of the most widely used tools. Specifically, a spatial branch-and-bound framework (Al-Khayyal et al. 1995, Audet et al. 2000, Linderoth 2005) is employed to search for the optimal solution. At each node of the branch-and-bound tree, a relaxation of the subproblem must be solved to provide a dual bound for pruning. The quality of this relaxation significantly impacts computational time, as tighter relaxations can substantially reduce search time. Current relaxation techniques primarily include McCormick inequalities (Misener and Floudas 2012), the reformulation-linearization technique (RLT) (Sherali and Adams 2013), SDP (Burer and Vandenbussche 2008), and various combinations thereof. However, these techniques often yield loose dual bounds, leading to a time-consuming branch-and-bound process. In this thesis, we introduce a novel approach to developing tighter relaxations for non-convex QCPs (Chapter 3).

The final challenging problem studied in this thesis is the planning of distributed energy resources in power grids, which is complicated by the significant uncertainties and complexities of distribution systems (Ehsan and Yang 2018, Oree et al. 2017). To ensure system reliability, we employ chance-constrained programs to seek highly likely feasible solutions while minimizing costs, resulting in a chance-constrained formulation of the planning problem. The traditional sample average approximation (SAA) is commonly used to represent uncertainties and reformulate chance-constrained programs into deterministic optimization problems (Pagnoncelli et al. 2009). However, SAA introduces additional binary variables to indicate whether a scenario sample is satisfied, adding computational complexity to an already challenging problem. To address this, we propose a new paradigm to improve computational tractability (Chapter 4).

1.2 Thesis Outline

In Chapter 2, we consider moment-based DRO, which provides a theoretical framework to integrate moment-based information from available data with optimal decision-making. Extensive studies have demonstrated the effectiveness of this framework in solving various industrial applications under uncertainties. Although moment-based DRO problems can be reformulated as SDPs that can be solved in polynomial time, solving high-dimensional SDPs is significantly challenging. More importantly, high-dimensional random parameters are generally involved in industrial applications, demanding efficient approaches to solve the high-dimensional SDPs in the context of moment-based DRO. Current approaches adopt the principal component analysis (PCA) to first reduce the dimensionality of random parameters using only the statistical information and then solve the subsequent low-dimensional approximation. We show that performing dimensionality reduction using the components with the largest variability may not produce a good optimal value from the subsequent PCA approximation and it can be even worse than using the components with the least variability. Thus, we integrate the dimensionality reduction with subsequent SDP problems and hence propose an optimized dimensionality reduction (ODR) approach for the moment-based DRO, aiming to drastically reduce the computational time of solving the SDP reformulations while maintaining the optimal solution of the original problem. We first derive an outer approximation under the ODR approach to provide a lower bound for the optimal value of the original problem, where the lower bound is nondecreasing in the reduced dimension m_1. We aim to choose a small m_1 to close the approximation gap between the derived lower bound and the original optimal value. To that end, we show that the rank of each SDP matrix with respect to an optimal solution of the original high-dimensional SDP reformulation is small, guiding us on how to optimize the dimensionality reduction. With this low-rank property, we observe that the derived lower bound can be close to the original optimal value but may not reach it. Nevertheless, it demonstrates that we are not far from closing the approximation gap and motivates us to derive an inner approximation to provide an upper bound for the optimal value of the original problem. More importantly, this upper bound reaches the original optimal value when the reduced dimension m_1 is small. Building on this significant result, we further derive an outer approximation to provide the second lower bound for the optimal value of the original problem, where the gap between the new lower bound and the original optimal value can be closed when the reduced dimension m_1 is small. The two outer and one inner approximations derived for the original problem are all low-dimensional SDPs and nonconvex with bilinear terms. We accordingly develop modified alternating direction method of multipliers (ADMM) algorithms to solve them efficiently and analyze the convergence property of the ADMM algorithms. Based on the near-optimal dimensionality reduction solution returned by the ADMM algorithms, we also explain how to recover the corresponding lower and upper bounds for the original optimal value. Finally, we demonstrate the effectiveness of our ODR approach in solving multiproduct newsvendor and production-transportation problems. We compare our ODR approach and algorithms with three benchmark approaches: the Mosek solver, the low-rank algorithm by Burer and Monteiro (2003), and existing PCA approximations by Cheramin et al. (2022). The numerical results show that our ODR approach significantly outperforms these benchmarks in computational time and solution quality. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. More importantly, unlike the existing approaches that become more computationally challenging when the dimension m of random parameters increases, our approach is not sensitive to m, demonstrating significant strength in solving large-scale practical problems. In addition, we provide insights into why our ODR approach performs better than the existing PCA approximations.

In Chapter 3, we consider a non-convex QCP optimizes a linear objective function subject to non-convex quadratic constraints. Despite its widespread applications in various industries and academia, finding an optimal solution for a non-convex QCP remains a challenging task. Currently, most QCP solvers utilize the spatial branch-and-bound algorithm to search for an optimal solution. One of the most important components of this algorithm is deriving efficient lower bounds, which help reduce the size of the branch-and-bound tree. In this chapter, we propose two mixed-integer linear programming (MILP) approximations to provide tighter lower bounds for a non-convex QCP. Initially, considering a non-convex QCP, we employ an eigenvalue-based decomposition to reformulate non-convex quadratic constraints. Specifically, we transform each non-convex quadratic constraint into a second-order cone (SOC) constraint $\|\mathbf{y}\| \leq t$ and the complement of an SOC constraint $\|\mathbf{z}\| \geq t$. Subsequently, we introduce a polyhedral approximation for $\|\mathbf{y}\| \leq t$. Based on the approximation of $\|\mathbf{y}\| \leq t$, we develop an approximation for $\|\mathbf{z}\| \geq t$. Combining these approximation results, we obtain a linear program with complementarity constraints (LPCC) approximation for the non-convex QCP. Furthermore, we establish that the feasible solution of the LPCC satisfies the "ϵ-relaxed" constraints of the non-convex QCP. More importantly, we demonstrate that the optimal value of the LPCC converges to the optimal value of the non-convex QCP asymptotically. Finally, we show the boundedness of the LPCC when the original non-convex QCP is bounded. As a result, we reformulate the LPCC into an MILP by utilizing this boundedness property. Motivated by the trade-off between the number of variables and the number of constraints in the context of extended formulations for approximations, we provide an alternative MILP approximation, where the asymptotic convergence property also holds. Compared to the first approximation, the second approximation typically introduces fewer variables but may result in more constraints when the approximation quality parameter ϵ is small. We conduct extensive computational experiments to validate the effectiveness of our proposed approximations over our randomly generated instances and the unitbox instances from the literature. The numerical results demonstrate that the proposed approximations yield smaller optimality gaps within shorter computational times than a state-of-the-art commercial solver. Meanwhile, our proposed approximations show a significant improvement in the optimality gap compared to the LP, SOCP, and SDP relaxations of the original non-convex QCP. We also apply our approximations to two problems: the joint decision and estimation problem and the two-trust-region subproblem. The numerical results highlight the significant advantages of our proposed approximations in terms of computational time and optimality gap.

In chapter 4, we focus on a planning problem that decides the locations and capacities of renewable distributed generation (RDG) and energy storage (ES) units considering significant uncertainties and complexities of distribution systems (e.g., ACOPF). To support such a decision-making problem, we develop a novel two-stage chance-constrained (TCC) model to ensure system reliability, minimize costs, and improve renewable energy penetration. One key feature of our model is that the chance constraint ensures that all the operational constraints are satisfied simultaneously with a high probability, leading to system reliability. We use two sampling techniques to reformulate our developed model, leading to the standard SAA formulation and our proposed partial sample average approximation (PSAA) formulation. The novelty of the PSAA formulation is that it introduces only continuous variables corresponding to the samples (as compared to integer variables in the SAA formulation) and uses historical data to improve its performance. Our extensive experiments show that the PSAA formulation performs better than the SAA formulation. The PSAA provides better locations and capacities of the RDG and ES units in a shorter time with a lower total cost and achieves a better desired probability of ensuring system feasibility than the SAA. The PSAA also reduces the optimality gap by more than 50% as compared to the SAA. We finally demonstrate the significance of ES units in reducing total costs and improving the power system balance.

In Chapter 5, we provide concluding remarks for this thesis.

Chapter 2

Optimized Dimensionality Reduction for Moment-based Distributionally Robust Optimization

2.1 Introduction

Distributionally robust optimization (DRO) is a modeling framework that integrates statistical information with traditional optimization methods (Scarf 1958, Delage and Ye 2010). Under this framework, one assumes that the underlying joint distribution of random parameters runs in a distributional ambiguity set inferred from given data or prior belief and then optimizes their decisions against the worst-case distribution within the set (see Rahimian and Mehrotra 2019 and Lin et al. 2022 for detailed review).

To solve different applications, researchers study the DRO under various ambiguity sets. The ambiguity set plays a crucial role in connecting statistical information and optimization modeling, providing a flexible framework to model uncertainties and incorporate partial information of random parameters (e.g., from historical data and prior belief) into the model. Moreover, the performance of DRO depends significantly on the distributional ambiguity set (Mohajerin Esfahani and Kuhn 2018, Chen et al. 2023). This chapter focuses on moment-based ambiguity sets, which include all distributions satisfying certain moment constraints. Examples of such constraints include restricting the exact mean and covariance matrix (Scarf 1958) and bounding the first and second moments (Ghaoui et al. 2003, Delage and Ye 2010). Moment-based DRO has been extensively studied and has a wide range of applications in industry, including but not limited to newsvendor problems (Gallego and Moon 1993, Yue et al. 2006, Natarajan et al. 2018), portfolio optimization (Ghaoui et al. 2003, Goldfarb and Iyengar 2003, Zymler et al. 2013, Rujeerapaiboon et al. 2016, Li 2018, Lotfi and Zenios 2018), knapsack problems (Cheng et al. 2014), transportation problems (Zhang et al. 2017a, Ghosal and Wiesemann 2020), reward-risk ratio optimization (Liu et al. 2017), scheduling problems (Shehadeh et al. 2020), and machine learning (Lanckriet et al. 2002, Farnia and Tse 2016).

As a moment-based DRO model can be reformulated as a semi-infinite program (Xu et al. 2018), three approaches are mainly used to solve such a reformulation: (i) the cutting plane/surface method (Gotoh and Konno 2002, Mehrotra and Papp 2014), by which a solution is first obtained by considering a subset of the ambiguity set and cuts are then added iteratively until converging to an optimal solution; (ii) the dual method (Delage and Ye 2010, Bertsimas et al. 2019b), by which the inner optimization problem (e.g., a minimization problem) is dualized and integrated with the outer optimization problem (e.g., a maximization problem); (iii) the analytical method (Scarf 1958, Popescu 2007), by which the worst-case distribution is obtained and its properties are analyzed. Among these methods, the dual method is the most popular. Most literature focuses on convex reformulations of different moment-based DRO problems, mainly including second-order cone programming (SOCP) (Ghaoui et al. 2003, Lotfi and Zenios 2018, Goldfarb and Iyengar 2003) and semidefinite programming (SDP) (Ghaoui et al. 2003, Delage and Ye 2010, Cheng et al. 2014).

While SOCPs can be solved efficiently, theoretically efficient algorithms (e.g., the interior-point methods) to solve SDPs impose substantial demands on computational time and memory resources (Vandenberghe and Boyd 1996, Helmberg 2002), particularly when addressing high-dimensional SDPs. Widely adopted commercial solvers (e.g., Mosek) exhibit prohibitively long computational times when solving high-dimensional SDPs, and the computational burden escalates considerably even as the problem dimension increases gradually (see our numerical results in Section 2.7). Thus, it is of great interest to study efficient algorithms for solving SDPs in the context of moment-based DRO. Besides the generic methods (e.g., the interior point methods), two types of algorithms can speed up solving SDP reformulations of moment-based DRO: low-rank SDP algorithms and dimensionality reduction methods. First, some studies develop efficient algorithms by exploiting the low-rank properties of SDP constraints (Burer and Monteiro 2003, Yurtsever et al. 2021). These algorithms rarely have theoretical guarantees but are practically efficient. Specifically, the existing studies may reformulate convex SDPs as non-convex problems and subsequently develop efficient algorithms to deliver high-quality solutions within reduced time frames (Lemon et al. 2016). Second, dimensionality reduction techniques represent data with important statistical information while omitting the trivial one. In the context of moment-based DRO, such techniques can be extended to reduce the dimension of random parameters and approximate the high-dimensional SDP reformulations using low-dimensional SDPs (Cheng et al. 2018, Cheramin et al. 2022), thereby reducing computational time significantly.

However, both the general SDP algorithms and existing dimensionality reduction methods may not perform well for moment-based DRO. The general SDP algorithm aims to solve general SDPs and may fail to consider the specific structure of the moment-based DRO models. The existing dimensionality reduction methods fail to consider the subsequent optimization problems when reducing the dimensionality space. For example, Cheng et al. (2018) and Cheramin et al. (2022) first use the PCA to choose the random parameters corresponding to the largest eigenvalues and then solve the low-dimensional SDP problem with the chosen random parameters. Such a sequential process may not provide an optimal solution of the original problem because the aim of leveraging data is to reduce the dimensionality space by focusing on only the statistical information, rather than optimizing the subsequent SDP problems. Therefore, in this chapter, we integrate the dimensionality reduction with subsequent SDP problems, leading to an *optimized dimensionality reduction (ODR) approach for moment-based DRO*. We summarize our contributions as follows:

- Unlike the PCA approximation approaches (Cheng et al. 2018, Cheramin et al. 2022) that first reduce dimensionality and then solve approximation problems, we integrate the dimensionality reduction with the subsequent optimization problems and provide an ODR approach. With the ODR approach, we develop two outer and one inner approximations for the original problem, leading to three low-dimensional SDP problems that can be solved efficiently.

- We prove the low-rank property of the original high-dimensional SDP reformulations of moment-based DRO problems. Specifically, we show that there exists an optimal solution such that the ranks of matrices in SDP reformulations are less than the number of SDP constraints plus one. Such a property helps our low-dimensional approximations achieve the original optimal value, closing the approximation gap.

- The low-dimensional SDP problems are nonconvex with bilinear terms and we develop modified Alternating Direction Method of Multipliers (ADMM) algorithms to solve them efficiently. We prove that any accumulation point of the sequence produced by the ADMM algorithm satisfies the first-order stationary conditions of the low-dimensional bilinear SDP problem. We apply the ODR approach and ADMM algorithms to solve multiproduct newsvendor and production-transportation problems. We compare our ODR approach with three benchmark approaches: the Mosek solver, low-rank algorithm (Burer and Monteiro 2003), and PCA approximations (Cheramin et al. 2022). The results demonstrate that our ODR approach significantly outperforms them in terms of computational time and solution quality. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. More importantly, our approach is not sensitive to the dimension m of random parameters, while the benchmark approaches perform much worse when m is larger.

Note that the ODR approach echoes the recently emerging framework that integrates machine learning (e.g., parameter estimation) with decision-making (Bertsimas and Kallus 2020, Bertsimas and Koduri 2022, Elmachtoub and Grigas 2022). More relevant applications of such a framework are recently studied. For instance, Ban and Rudin (2019) and Zhang et al. (2023) integrate feature data within the newsvendor problem; Liu et al. (2021) integrate travel-time predictors with order-assignment optimization to provide last-mile delivery services; Kallus and Mao (2023) propose a new random forest algorithm that considers the downstream optimization problem; Zhu et al. (2022) develop a joint estimation and robustness optimization framework; Qi et al. (2023) and Ho-Nguyen and Kılınç-Karzan (2022) provide an end-to-end framework to integrate prediction and optimization. Unlike the above applications, we integrate dimensionality reduction with optimization in this chapter (Jiang et al. 2023), which is recently followed by He and Mak (2023). He and Mak (2023) integrate the PCA with a subsequent stochastic program and provide a distributionally robust bound for the error between the objective values of the original and integrated problems. The integrated approach in He and Mak (2023) involves solving nonconvex and high-dimensional SDPs and may not reduce the error to zero, while our approach solves low-dimensional SDPs and can achieve the optimal value of the original moment-based DRO problem, thereby offering guidance on selecting the reduced dimension for practical applicability.

The remainder of this chapter is organized as follows. Section 2.2 provides the SDP reformulation of moment-based DRO problems and illustrates the disadvantages of the PCA approximation approaches (Cheng et al. 2018, Cheramin et al. 2022). In Section 2.3, we propose the first outer approximation under the ODR approach, leading to a lower bound for the original problem, and are then motivated to develop the low-rank property of the original high-dimensional SDP reformulation, aiming to find a small reduced dimension to close the approximation gap. In Sections 2.4 and 2.5, motivated by the results in Section 2.3, we provide an inner approximation and a second outer approximation for the original problem, respectively, and both of them can achieve the original optimal value. Section 2.6 develops efficient algorithms to solve the above three approximations, all of which are low-dimensional bilinear SDP problems. In Section 2.7, we conduct extensive numerical experiments on multiproduct newsvendor and production-transportation problems. Section 2.8 concludes the chapter. All proofs are presented in the Appendix if not specified.

Notation

We use non-bold symbols to denote scalar values, e.g., s and γ_1, and bold symbols to denote vectors, e.g., $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $\mathbf{q}$. Similarly, matrices are represented by bold capital symbols, e.g., $\mathbf{A}$ and $\boldsymbol{\Sigma}$, and the size of a matrix is indicated by $r \times c$, where r and c indicate the numbers of rows and columns, respectively. Italic subscripts indicate indices, e.g., S_k, while non-italic ones represent simplified specifications, e.g., $\mathbf{Q}_{\mathrm{r}}$. We use $\mathbb{E}_{\mathbb{P}}[\cdot]$ to represent the expectation over distribution $\mathbb{P}$ and use " $\bullet$ " to denote the inner product defined by $\mathbf{A} \bullet \mathbf{B} = \sum_{i,j} A_{ij} B_{ij}$, where $\mathbf{A}$ and $\mathbf{B}$ are two conformal matrices. For any matrix $\mathbf{M}$, we use $\mathbf{M} \succeq 0$ (resp. $\mathbf{M} \succ 0$) to indicate that it is positive semi-definite (PSD) (resp. positive definite). Symbols $\|\cdot\|_1$ and $\|\cdot\|_2$ denote L1-Norm and L2-Norm, respectively. For any integer number $n \geq 1$, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. The identity matrix of size m is denoted by $\mathbf{I}_m$. Symbols $\mathbf{0}_m$ and $\mathbf{0}_{r\times c}$ represent a zero vector of size m and a zero matrix of size $r \times c$, respectively. Symbols $\mathbf{1}_m$ and $\mathbf{1}_{r\times c}$ represent a one vector of size m and a one matrix of size $r \times c$, respectively. We use $\mathbb{I}(\cdot)$ to denote the indicator function, which takes 1 if all the conditions encompassed in $(\cdot)$ are satisfied and takes 0 otherwise.

2.2 SDP Reformulation

Given the distribution $\mathbb{P}$ of a random vector $\boldsymbol{\xi} \in \mathbb{R}^m$, the following stochastic programming (SP) formulation seeks an $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ to minimize the expectation of an objective function $f(\mathbf{x}, \boldsymbol{\xi})$:

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}} \left[f\left(\mathbf{x}, \boldsymbol{\xi}\right) \right]. \tag{2.1}$$

As the distribution $\mathbb{P}$ is often unknown, we assume that $\mathbb{P}$ belongs to a distributional ambiguity set $\mathcal{D}_{\mathrm{M0}}$ constructed by statistical information estimated from historical data, and then minimize $f(\mathbf{x}, \boldsymbol{\xi})$ against the worst-case distribution instead. It leads to the following DRO formulation:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P} \in \mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}} \left[f\left(\mathbf{x}, \boldsymbol{\xi}\right) \right]. \tag{2.2}$$

We consider moment-based information (Delage and Ye 2010) is included in $\mathcal{D}_{\mathrm{M0}}$ as follows:

$$\mathcal{D}_{\mathrm{M0}}\left(\mathcal{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma_1, \gamma_2\right) = \left\{ \mathbb{P} \;\middle|\; \begin{array}{l} \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{S}\right) = 1, \quad \left(\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\xi}\right] - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\xi}\right] - \boldsymbol{\mu}\right) \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{\xi} - \boldsymbol{\mu}\right)\left(\boldsymbol{\xi} - \boldsymbol{\mu}\right)^\top\right] \preceq \gamma_2 \boldsymbol{\Sigma} \end{array} \right\},$$

which describes that (i) the support of $\boldsymbol{\xi}$ is $\mathcal{S}$, (ii) the mean of $\boldsymbol{\xi}$ lies in an ellipsoid of size γ_1 centered at $\boldsymbol{\mu}$, and (iii) the covariance of $\boldsymbol{\xi}$ is bounded from above by $\gamma_2 \boldsymbol{\Sigma}$, with $\gamma_1 \geq 0$, $\gamma_2 \geq 1$, and $\boldsymbol{\Sigma} \succ 0$. Note that $\boldsymbol{\mu}$ is the sample mean and $\boldsymbol{\Sigma}$ is the sample covariance matrix, which can be estimated from the data. Meanwhile, the parameters γ_1 and γ_2 provide natural means of quantifying our confidence in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, and So (2011) proposes a way to select them. We perform eigenvalue decomposition on the covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\boldsymbol{\Sigma} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top = \mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}} \left(\mathbf{U} \boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top,$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix and $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ is a diagonal matrix. Without loss of generality, we assume that the diagonal elements of $\mathbf{\Lambda}$ are arranged in a nonincreasing order. By letting $\boldsymbol{\xi} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}$, we can reformulate Problem (2.2) as:

$$\Theta_{\mathrm{M}}(m) := \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right], \tag{2.3}$$

where

$$\mathcal{D}_{\mathrm{M}}\left(\mathcal{S}_{\mathrm{I}}, \gamma_1, \gamma_2\right) = \left\{ \mathbb{P}_{\mathrm{I}} \;\middle|\; \begin{array}{l} \mathbb{P}_{\mathrm{I}}\left(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}\right) = 1, \quad \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^{\top}\right] \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}\right] \preceq \gamma_2 \mathbf{I}_m \end{array} \right\},$$

with $\mathcal{S}_{\mathrm{I}} := \{\boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m \mid \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu} \in \mathcal{S}\}$. Similar to Cheng et al. (2018) and Cheramin et al. (2022), we make the following assumption throughout the chapter.

Assumption 1. *Function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex in $\boldsymbol{\xi}$, i.e., $f(\mathbf{x}, \boldsymbol{\xi}) = \max_{k=1}^{K} \{y_k^0(\mathbf{x}) + y_k(\mathbf{x})^{\top}\boldsymbol{\xi}\}$ with $y_k(\mathbf{x}) = (y_k^1(\mathbf{x}), \ldots, y_k^m(\mathbf{x}))^{\top}$ and $y_k^0(\mathbf{x})$ affine in $\mathbf{x}$ for any $k \in [K]$, and $\mathcal{S}$ is polyhedral, i.e., $\mathcal{S} = \{\boldsymbol{\xi} \mid \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$ with $\mathbf{A} \in \mathbb{R}^{l \times m}$ and $\mathbf{b} \in \mathbb{R}^l$, with at least one interior point.*

Note that Assumption 1 specifies a wide range of objective functions and support of uncertainty (Cheramin et al. 2022). Meanwhile, it helps us reduce Problem (2.3) to an SDP formulation.

Proposition 1 (Cheramin et al. 2022)**.** *Under Assumption 1, Problem (2.3) has the same optimal value as the following SDP formulation:*

$$\Theta_{\mathrm{M}}(m) = \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}, \mathbf{Q}}} \quad \phi(m, s, \mathbf{q}, \mathbf{Q}) \tag{2.4a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right)^{\top} \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{2.4b}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}_+^l, \ \forall k \in [K], \ \mathbf{x} \in \mathcal{X}, \tag{2.4c}$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$, $\mathbf{q} \in \mathbb{R}^m$, $\mathbf{Q} \in \mathbb{R}^{m\times m}$, $\phi(m, s, \mathbf{q}, \mathbf{Q}) := s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1}\|\mathbf{q}\|_2$, $\chi(k, \mathbf{x}, s, \boldsymbol{\lambda}) := s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}^\top \mathbf{A}\boldsymbol{\mu}$, *and* $\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}) := \mathbf{A}^\top \boldsymbol{\lambda} - y_k(\mathbf{x})$.

Note that the functions $\phi(m, s, \mathbf{q}, \mathbf{Q})$, $\chi(k, \mathbf{x}, s, \boldsymbol{\lambda})$, and $\boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda})$ will also be used in the remainder of this chapter to simplify other SDP formulations. Although Problem (2.4) is a convex program when $\mathbf{x}$ is given, it can be difficult to solve because a large m leads to high-dimensional SDP constraints at size $m + 1$. As such SDP constraints originate from the covariance matrix $\boldsymbol{\Sigma}$, early attempts in Cheng et al. (2018) and Cheramin et al. (2022) exploit the statistical information $\boldsymbol{\Sigma}$ to address the computational challenge while maintaining solution quality. Specifically, they adopt the PCA, a dimensionality reduction method commonly used in statistical learning, to capture the dominant variability of $\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}$ by maintaining the first $m_1 (\leq m)$ components of $\boldsymbol{\xi}_{\mathrm{I}}$ and fixing its other components at 0; that is,

$$\boldsymbol{\xi} \approx \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left[\boldsymbol{\xi}_{\mathrm{r}}; \mathbf{0}_{m-m_1}\right] + \boldsymbol{\mu} = \mathbf{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}, \tag{2.5}$$

where $\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1}$, and $\mathbf{U}_{m\times m_1} \in \mathbb{R}^{m\times m_1}$ and $\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}} \in \mathbb{R}^{m_1\times m_1}$ are upper-left submatrices of $\mathbf{U}$ and $\boldsymbol{\Lambda}$, respectively. That is, the m_1 components of $\boldsymbol{\xi}_{\mathrm{I}}$ corresponding to the largest eigenvalues are maintained as uncertain and the other components are fixed at their means. With a lower-dimensional random vector $\boldsymbol{\xi}_{\mathrm{r}}$, we can have a relaxation of Problem (2.3):

$$\Theta_{\mathrm{M}}(m_1) := \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}}\in\mathcal{D}_{\mathrm{L}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[f\left(\mathbf{x}, \mathbf{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}\right)\right], \tag{2.6a}$$

where

$$\mathcal{D}_{\mathrm{L}}\left(\mathcal{S}_{\mathrm{r}}, \gamma_1, \gamma_2\right) = \left\{ \mathbb{P}_{\mathrm{r}} \;\middle|\; \begin{array}{l} \mathbb{P}_{\mathrm{r}}\left(\boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}\right) = 1, \quad \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^\top\right] \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\right] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\xi}_{\mathrm{r}}^\top\right] \preceq \gamma_2 \mathbf{I}_{m_1} \end{array} \right\} \tag{2.6b}$$

with

$$\mathcal{S}_{\mathrm{r}} := \left\{ \boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \;\middle|\; \mathbf{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S} \right\}. \tag{2.6c}$$

Meanwhile, the corresponding SDP formulation of Problem (2.6) has SDP constraints with smaller size at m_1+1 and can be solved more efficiently than Problem (2.4), leading to an efficient "*PCA approximation.*" Specifically, Cheramin et al. (2022) show that the following PCA approximation

$$\Theta_{\mathrm{M}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q}_{\mathrm{r}},\mathbf{Q}_{\mathrm{r}}}} \quad \phi(m_1, s, \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}) \tag{2.7a}$$

$$\text{s.t.} \begin{bmatrix} \chi(k,\mathbf{x},s,\boldsymbol{\lambda}_k) & \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)\right)^{\top} \\ \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}_{m\times m_1}\boldsymbol{\Lambda}_{m_1}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{2.7b}$$

$$\boldsymbol{\lambda}_k \in \mathbb{R}^l_+, \ \forall k \in [K], \mathbf{x} \in \mathcal{X}, \tag{2.7c}$$

where $\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$, $\mathbf{q}_{\mathrm{r}} \in \mathbb{R}^{m_1}$, and $\mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1}$, provides a *lower bound* for the optimal value of Problem (2.3) (i.e., Problem (2.4)). The PCA approximation that leads to an *upper bound* for the optimal value of Problem (2.3) can be similarly derived. Note that in the PCA approximation, we need to perform eigenvalue decomposition on the covariance matrix $\boldsymbol{\Sigma}$, which can be time-consuming when the dimension of $\boldsymbol{\xi}$ is high. However, the time for eigenvalue decomposition is negligible compared with that of solving the SDP problem. Therefore, in this thesis, we mainly focus on the dimensionality reduction of the SDP problem. Hereafter, we call the problem whose optimal value is a lower bound of the original Problem (2.3) as an *outer approximation.* In contrast, the problem generating an upper bound is called an *inner approximation* of Problem (2.3).

However, relying on only the statistical information (i.e., dominant variability) to choose the components and reducing the high-dimensional uncertainty space may not lead to the best approximation performance. Although Cheramin et al. (2022) provide a performance guarantee (Propositions 3 and 6 in Cheramin et al. (2022)) to bound the gap between the original and approximated objective values, it is difficult to close the gap when reducing the dimensionality of $\boldsymbol{\xi}_{\mathrm{I}}$. Such a difficulty is not surprising because maintaining only the largest statistical variability in the PCA approximations does not capture the optimality conditions of the original problems (e.g., Problem (2.3)). We provide an example as follows to illustrate that choosing the components of $\boldsymbol{\xi}_{\mathrm{I}}$ corresponding to the largest eigenvalues can be even worse than choosing the components corresponding to the least eigenvalues.

Example 1. *Given* $\mathbf{x} \in \mathcal{X}$*, we consider the* $CVaR_{1-\alpha}$ *of a cost function* $g(\mathbf{x}, \boldsymbol{\xi})$ *formulated as the following optimization problem (Rockafellar and Uryasev 2000):*

$$\min_{t \in \mathbb{R}} \; t + \frac{1}{\alpha} \mathbb{E}_{\mathbb{P}} \left[g(\mathbf{x}, \boldsymbol{\xi}) - t \right]^{+}, \tag{2.8}$$

where $\alpha \in (0,1)$ *is a risk tolerance level and function* $[\cdot]^{+} := \max\{0, \cdot\}$*. For brevity, we let* $g(\mathbf{x}, \boldsymbol{\xi}) = \mathbf{x}^{\top}\boldsymbol{\xi}$*,* $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^m_+ \mid \sum_{i=1}^m x_i = 1\}$*,* $\mathcal{D} = \{\mathbb{P} \mid \mathbb{P}(\boldsymbol{\xi} \in \mathcal{S}) = 1,\ \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})^{\top}] \preceq \boldsymbol{\Sigma}\}$*,* $\mathcal{S}$ *is compact, and* $\boldsymbol{\mu}$ *is in the interior of* $\mathcal{S}$*. We reformulate the distributionally robust counterpart of Problem (2.8) in Appendix A.2.1 and obtain Problem (A.2). Let* $\alpha = 0.05$*,* $\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^3 \mid 0 \le \xi_1 \le 8, 1 \le \xi_2 \le 12, 2 \le \xi_3 \le 16\}$*,* $\boldsymbol{\mu} = [1,2,3]$*,* $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 3 & 0.15 \\ 0.1 & 0.15 & 2 \end{bmatrix}$ *with eigenvalues* 3.044*,* 1.983*, and* 0.973*. Solving Problem (A.2) gives the optimal value* 5.021 *with* $x_1 = 0.719$*,* $x_2 = 0.135$*,* $x_3 = 0.145$*, and* $t = 3.129$*. Following Cheng et al. (2018) and Cheramin et al. (2022) to perform PCA approximation over Problem (A.2) by capturing only one of the three components in* $\boldsymbol{\xi}$*, we observe the following:*

- *Choosing the component corresponding to the largest eigenvalue* 3.044*, the PCA approximation gives the optimal value at* 1.788 *with* $x_1 = 1$*,* $x_2 = 0$*,* $x_3 = 0$*, and* $t = 1.373$*.*
- *Choosing the component corresponding to the second largest eigenvalue* 1.983*, the PCA approximation gives the optimal value at* 1.3 *with* $x_1 = 0.7$*,* $x_2 = 0.3$*,* $x_3 = 0$*, and* $t = 1.3$*.*
- *Choosing the component corresponding to the smallest eigenvalue* 0.973*, the PCA*

approximation gives the optimal value at 1.915 *with* $x_1 = 0.085$, $x_2 = 0.915$, $x_3 = 0$, *and* $t = 1.915$.

Example 1 shows that performing dimensionality reduction (i.e., from $\boldsymbol{\xi}$ to $\boldsymbol{\xi}_\mathrm{r}$) using the components with the largest variability may not produce a good optimal value from the *subsequent* PCA approximation (i.e., an SDP) and it can be even worse than using the component with the smallest variability. To solve this issue, we integrate the dimensionality reduction with the subsequent approximation in the following sections, leading to an *optimized dimensionality reduction (ODR) approach*. Correspondingly, we obtain efficient lower and upper bounds in the following Sections 2.3–2.5 and more importantly, the bounds can achieve the optimal value of the original Problem (2.3).

2.3 Lower Bound

We extend the dimensionality reduction method (i.e., PCA) in (2.5) by introducing a decision variable $\mathbf{B} \in \mathcal{B}_{m_1} := \{\mathbf{B} \in \mathbb{R}^{m \times m_1} \mid \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}\} \subseteq \mathbb{R}^{m \times m_1}$ such that

$$\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu} \approx \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_\mathrm{r} + \boldsymbol{\mu}, \tag{2.9}$$

where $\mathbf{B}$ will be optimized in the subsequent PCA approximation, i.e., optimized dimensionality reduction. By (2.9), we project $\boldsymbol{\xi}_\mathrm{I}$ onto a subspace of $\mathbb{R}^{m \times m}$ spanned by the columns of $\mathbf{B} \in \mathcal{B}_{m_1}$ and approximate $\boldsymbol{\xi}_\mathrm{I}$ by the projection $\mathbf{B}\boldsymbol{\xi}_\mathrm{r}$, instead of considering only the random variables corresponding to the largest eigenvalues. When $\mathbf{B} = \begin{bmatrix} \mathbf{I}_{m_1} \\ \mathbf{0}_{(m-m_1)\times m_1} \end{bmatrix}$, (2.9) reduces to (2.5). Therefore, we would like to choose a good (even an optimal) $\mathbf{B}$ to obtain a better lower bound for Problem (2.3) than Problem (2.7). Unlike the existing PCA approach that first reduces the dimension of the uncertainty space and then provides approximations, our ODR method innovatively integrates dimensionality reduction with the subsequent optimization problems. Such an integrated framework deviates from the traditional dimensionality reduction method like PCA because we do not predetermine a low-dimensional space to consider in the subsequent optimization problem. Instead, we linearly map the high-dimensional uncertainty space to a low-dimensional space while such a mapping relationship (represented by the decision $\mathbf{B}$) is carefully optimized together with the subsequent optimization problems.

Given any $m_1 \in [m]$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, we obtain a relaxation of Problem (2.3) by extending Problem (2.6). If the relaxation provides a lower bound for the optimal value of Problem (2.3), then we may choose the best $\mathbf{B} \in \mathcal{B}_{m_1}$ such that we obtain the largest possible lower bound. Thus, we build the following *integrated dimensionality reduction and optimization* problem:

$$\Theta_{\mathrm{L}}(m_1) = \max_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}} \in \mathcal{D}_{\mathrm{L}}} \mathbb{E}_{\mathbb{P}_{\mathrm{r}}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \right) \right], \tag{2.10}$$

where $\mathcal{D}_{\mathrm{L}}$ is defined in (2.6b) with

$$\mathcal{S}_{\mathrm{r}} := \left\{ \boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \,\middle|\, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S} \right\}. \tag{2.11}$$

We will show that Problem (2.10) provides a lower bound for Problem (2.3) (see Theorem 1). Before presenting this theorem, we prepare the following two lemmas.

Lemma 1. *When $\mathbf{B} \in \mathbb{R}^{m \times m_1}$, the following three constraints are equivalent: (i) $\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0$, (ii) $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$, and (iii) $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$.*

Lemma 1 shows that both $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$ and $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$ can be reformulated as an SDP constraint $\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0$. Although this SDP constraint has a high dimension at $m+m_1$, it is very sparse and usually does not create additional computational challenges.

Lemma 2. *For any matrix $\mathbf{V} \in \mathbb{R}^{m \times n}$ and symmetric matrices $\mathbf{X} \in \mathbb{R}^{m \times m}$ and $\mathbf{Y} \in \mathbb{R}^{m \times m}$, we have: (i) If $\mathbf{X} \succeq \mathbf{Y}$, then $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$; (ii) If $n = m$ and $\mathbf{V}$ is invertible, then $\mathbf{X} \succeq \mathbf{Y}$ is equivalent to $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$.*

Lemma 2 shows that a PSD matrix (e.g., $\mathbf{X}-\mathbf{Y}$) remains PSD if it is pre-multiplied by an arbitrary matrix with appropriate dimensions (e.g., $\mathbf{V}^\top$) and post-multiplied by this arbitrary matrix's transpose (e.g., $\mathbf{V}$). Furthermore, if this arbitrary matrix is invertible, then the original PSD matrix is equivalent to the matrix after the pre-multiplication and post-multiplication. With Lemmas 1 and 2, we are now ready to present the following theorem.

Theorem 1. *The following three conclusions hold: (i) Problem* (2.10) *provides a lower bound for the optimal value of Problem* (2.3)*, i.e.,* $\Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}(m)$ *for any* $m_1 \leq m$*; (ii) the optimal value of Problem* (2.10) *is nondecreasing in* m_1*, i.e.,* $\Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{L}}(m_2)$ *for any* $m_1 < m_2 \leq m$*; and (iii) when* $m_1 = m$*, Problem* (2.3) *and Problem* (2.10) *have the same optimal value, i.e.,* $\Theta_{\mathrm{L}}(m) = \Theta_{\mathrm{M}}(m)$*.*

Theorem 1 shows that we obtain a lower bound for the optimal value of Problem (2.3) when reducing the dimensionality space of $\boldsymbol{\xi}_{\mathrm{I}}$ *while optimizing the choice of* $\mathbf{B} \in \mathcal{B}_{m_1}$ in Problem (2.10). When the reduced dimensionality (i.e., m_1) is higher, we obtain a better lower bound. We maintain the optimal value of Problem (2.3) if the dimensionality space is not reduced (i.e., $m_1 = m$). Note that the conclusions in Theorem 1 are similar to those in Theorem 2 in Cheramin et al. (2022), both demonstrating the validity of dimensionality reduction in solving the moment-based DRO problems. However, here by optimizing the choice of $\mathbf{B} \in \mathcal{B}_{m_1}$, Problem (2.10) provides a better lower bound than Problem (2.6) (i.e., the PCA approximation in Cheramin et al. 2022) does because the latter problem is a special case of the former problem. More importantly, we may expect to close the gap between $\Theta_{\mathrm{L}}(m_1)$ and $\Theta_{\mathrm{M}}(m)$ when we choose a small m_1. To that end, we follow the PCA approximation (2.7) to reformulate Problem (2.10) as the following SDP formulation:

$$\Theta_{\mathrm{L}}(m_1) = \max_{\mathbf{B} \in \mathcal{B}_{m_1}} \underline{\Theta}(m_1, \mathbf{B}), \tag{2.12}$$

where

$$\underline{\Theta}(m_1, \mathbf{B}) := \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \\ \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}}} \quad \phi(m_1, s, \mathbf{q}_{\mathrm{r}}, \mathbf{Q}_{\mathrm{r}}) \tag{2.13a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right)^\top \\ \frac{1}{2}\left(\mathbf{q}_{\mathrm{r}} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k)\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{2.13b}$$

$$\mathbf{x} \in \mathcal{X}; \;\; \hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}, \;\; \boldsymbol{\lambda}_k \in \mathbb{R}^l_+, \;\; \forall k \in [K]; \;\; \mathbf{q}_{\mathrm{r}} \in \mathbb{R}^{m_1}; \;\; \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1 \times m_1}. \tag{2.13c}$$

Now we would like to find an $m_1 < m$ such that $\Theta_{\mathrm{L}}(m_1)$ in Problem (2.12) is close (even equal) to $\Theta_{\mathrm{M}}(m)$ in Problem (2.4). Note that if $\Theta_{\mathrm{L}}(m_1) = \Theta_{\mathrm{M}}(m)$, then comparing the SDP constraints between (2.4) and (2.12) shows that the rank of $\mathbf{Q}$ in the optimal solution of Problem (2.4) can be smaller than m. Specifically, we are motivated to explore the low-rank property of Problem (2.4) and obtain the following significant conclusion.

Theorem 2. *Consider $K < m$ and any optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$ of Problem (2.4) with $S_k = s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top}\mathbf{b} - y_k(\mathbf{x}^*)^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top}\mathbf{A}\boldsymbol{\mu}$ for any $k \in [K]$. We can always construct another optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}', \mathbf{Q}')$ of Problem (2.4) such that $\mathrm{rank}(\mathbf{Q}') \leq K$, $\mathbf{q}' = \mathbf{V}\boldsymbol{\delta}$, $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, and $(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)) = \mathbf{V}\boldsymbol{\nu}_k$ for any $k \in [K]$, where $\mathbf{Y}_{11} \in \mathbb{R}^{K\times K}$, $\mathbf{Y}_{11} \succeq 0$, $\mathbf{V} = [\mathbf{v}_k, \; \forall k \in [K]] \in \mathbb{R}^{m\times K}$ with orthonormal vectors $\mathbf{v}_k \in \mathbb{R}^m$, $\boldsymbol{\delta} \in \mathbb{R}^K$, and $\boldsymbol{\nu}_k \in \mathbb{R}^K$ for any $k \in [K]$ depend on the optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$.*

Proof. Proof. Note that the optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}^*, \mathbf{Q}^*)$ of Problem (2.4) leads to the optimal value $s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\,\|\mathbf{q}^*\|_2$. Based on this optimal solution, we construct a feasible solution of Problem (2.4), denoted by $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ such that $\mathbf{x}' = \mathbf{x}^*$, $s' = s^*$, and $\hat{\boldsymbol{\lambda}}' = \hat{\boldsymbol{\lambda}}^*$.

Now we construct the values of $\mathbf{q}'$ and $\mathbf{Q}'$. By constraints (2.4b), we have

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k\left(\mathbf{x}^*\right)\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k\left(\mathbf{x}^*\right)\right)\right) & \mathbf{Q}^* \end{bmatrix} \succeq 0,$$
$$\forall k \in [K]. \tag{2.14}$$

We can equivalently rewrite (2.14) as

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top, \forall k \in [K]; \mathbf{Q}^* \succeq 0. \tag{2.15}$$

Note that, if $S_k > 0$ for any $k \in [K]$, then (2.14) is equivalent to (2.15) by Schur complement; otherwise, when $S_k = 0$ for some $k \in [K]$, we have $(1/2)(\mathbf{q}^* + (\mathbf{U}\boldsymbol{\Lambda}^{1/2})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))) = \mathbf{0}_m$ by the definition of a PSD matrix. Thus, (2.14) is equivalent to $\mathbf{Q}^* \succeq 0$, i.e., (2.15).

Note that $K < m$. Thus, through the Gram–Schmidt process, we can always construct K orthonormal vectors $\mathbf{v}_k \in \mathbb{R}^m,\ \forall k \in [K]$, and K real vectors $\boldsymbol{\kappa}_k \in \mathbb{R}^K,\ \forall k \in [K]$, such that

$$\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right) = \mathbf{V}\boldsymbol{\kappa}_k,\ \forall k \in [K], \tag{2.16}$$

$\mathbf{V} = [\mathbf{v}_k,\ \forall k \in [K]] \in \mathbb{R}^{m\times K}$. We further extend $\mathbf{V}$ to $[\mathbf{V}, \bar{\mathbf{V}}] \in \mathbb{R}^{m\times m}$ with $\bar{\mathbf{V}} \in \mathbb{R}^{m\times(m-K)}$ such that all the column vectors of $[\mathbf{V}, \bar{\mathbf{V}}]$ can span the space of $\mathbb{R}^m$. As $\mathbf{q}^* \in \mathbb{R}^m$, we can find $\boldsymbol{\kappa}_0 \in \mathbb{R}^K$ and $\bar{\boldsymbol{\kappa}}_0 \in \mathbb{R}^{m-K}$ such that

$$\mathbf{q}^* = \mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0. \tag{2.17}$$

As $\mathbf{Q}^* \in \mathbb{R}^{m\times m}$, we can then decompose $\mathbf{Q}^*$ as

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{V} & \bar{\mathbf{V}} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}^\top \\ \bar{\mathbf{V}}^\top \end{bmatrix} = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top, \tag{2.18}$$

where $\mathbf{Y}_{11} \in \mathbb{R}^{K\times K}$, $\mathbf{Y}_{12} \in \mathbb{R}^{K\times(m-K)}$, $\mathbf{Y}_{21} \in \mathbb{R}^{(m-K)\times K}$, and $\mathbf{Y}_{22} \in \mathbb{R}^{(m-K)\times(m-K)}$. As $\mathbf{Q}^* \succeq 0$, we have $\begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{v} & \bar{\mathbf{v}} \end{bmatrix}^{-1} \mathbf{Q}^* \begin{bmatrix} \mathbf{V}^\top \\ \bar{\mathbf{V}}^\top \end{bmatrix}^{-1} \succeq 0$ by Lemma 2. By (2.15), (2.16), and (2.17), we have

$$4S_k\mathbf{Q}^* \succeq \left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)\left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)^\top, \ \forall k \in [K]. \tag{2.19}$$

By (2.18) and (2.19), we have

$$\begin{aligned} 4S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top\right) \succeq& \\ \left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)&\left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k\right)^\top, \forall k \in [K]. \end{aligned}$$

By Lemma 2, we further have

$$\begin{aligned} &4S_k\mathbf{V}^\top(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top + \bar{\mathbf{V}}\mathbf{Y}_{21}\mathbf{V}^\top + \mathbf{V}\mathbf{Y}_{12}\bar{\mathbf{V}}^\top + \bar{\mathbf{V}}\mathbf{Y}_{22}\bar{\mathbf{V}}^\top)\mathbf{V} \\ &\qquad \succeq \mathbf{V}^\top(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k)(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0 + \mathbf{V}\boldsymbol{\kappa}_k)^\top\mathbf{V}, \ \forall k \in [K]. \end{aligned} \tag{2.20}$$

Because $\mathbf{V}^\top\bar{\mathbf{V}} = \mathbf{0}$, $\bar{\mathbf{V}}^\top\mathbf{V} = \mathbf{0}$, and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, constraints (2.20) become

$$4S_k\mathbf{Y}_{11} \succeq (\boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_k)(\boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_k)^\top, \ \forall k \in [K]. \tag{2.21}$$

Now we let $\mathbf{q}' = \mathbf{V}\boldsymbol{\kappa}_0$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$. By (2.21) and Lemma 2, we have

$$4S_k\mathbf{Q}' = 4S_k\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \succeq (\mathbf{V}\boldsymbol{\kappa}_0 + \mathbf{V}\boldsymbol{\kappa}_k)(\mathbf{V}\boldsymbol{\kappa}_0 + \mathbf{V}\boldsymbol{\kappa}_k)^\top$$

$$= \left(\mathbf{q}' + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) \left(\mathbf{q}' + \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top, \forall k \in [K]. \tag{2.22}$$

Comparing (2.4b) and (2.22), we have $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (2.4) and the corresponding objective value is

$$s' + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\left\|\mathbf{q}'\right\|_2 \geq s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1}\left\|\mathbf{q}^*\right\|_2, \tag{2.23}$$

where the inequality holds because $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (2.4) and Problem (2.4) is a minimization problem. Note that

$$\begin{aligned}
\mathbf{I}_m \bullet \mathbf{Q}^* &= tr(\mathbf{Q}^*) = tr\left(\begin{bmatrix}\mathbf{V} & \bar{\mathbf{V}}\end{bmatrix}\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}\right)\\
&= tr\left(\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}\begin{bmatrix}\mathbf{V} & \bar{\mathbf{V}}\end{bmatrix}\right) = tr\left(\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix}\right)\\
&= \mathbf{I}_K \bullet \mathbf{Y}_{11} + \mathbf{I}_{m-K} \bullet \mathbf{Y}_{22} \geq \mathbf{I}_K \bullet \mathbf{Y}_{11} = tr(\mathbf{Y}_{11}) = tr(\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V})\\
&= tr(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) = tr(\mathbf{Q}'),
\end{aligned}$$

where the first equality holds by the definition of a matrix's trace, the second equality holds by (2.18), the third equality holds by the cyclic property of a matrix's trace, the fourth equality holds because $\begin{bmatrix}\mathbf{V}^\top\\ \bar{\mathbf{V}}^\top\end{bmatrix}\begin{bmatrix}\mathbf{V} & \bar{\mathbf{V}}\end{bmatrix} = \mathbf{I}_m$, and the first inequality holds because $\begin{bmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12}\\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{bmatrix} \succeq 0$ and accordingly $\mathbf{I}_{m-K} \bullet \mathbf{Y}_{22} \geq 0$. Meanwhile,

$$\begin{aligned}
\left\|\mathbf{q}^*\right\|_2^2 &= (\mathbf{q}^*)^\top \mathbf{q}^* = \left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0\right)^\top \left(\mathbf{V}\boldsymbol{\kappa}_0 + \bar{\mathbf{V}}\bar{\boldsymbol{\kappa}}_0\right) = \left(\boldsymbol{\kappa}_0^\top\boldsymbol{\kappa}_0 + \bar{\boldsymbol{\kappa}}_0^\top\bar{\boldsymbol{\kappa}}_0\right)\\
&\geq \boldsymbol{\kappa}_0^\top\boldsymbol{\kappa}_0 = (\mathbf{V}\boldsymbol{\kappa}_0)^\top(\mathbf{V}\boldsymbol{\kappa}_0) = (\mathbf{q}')^\top\mathbf{q}' = \left\|\mathbf{q}'\right\|_2^2,
\end{aligned}$$

where the second equality holds by (2.17), the third equality holds because $\mathbf{V}^\top\bar{\mathbf{V}} = \mathbf{0}$, $\bar{\mathbf{V}}^\top\mathbf{V} = \mathbf{0}$, $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, $\bar{\mathbf{V}}^\top\bar{\mathbf{V}} = \mathbf{I}_{m-K}$, and the first inequality holds because $\bar{\boldsymbol{\kappa}}_0^\top\bar{\boldsymbol{\kappa}}_0 \geq$

0. Thus, we have

$$s' + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1} \left\| \mathbf{q}' \right\|_2 \le s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1} \left\| \mathbf{q}^* \right\|_2 . \tag{2.24}$$

Combining (2.23) and (2.24) leads to

$$s' + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1} \left\| \mathbf{q}' \right\|_2 = s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}^* + \sqrt{\gamma_1} \left\| \mathbf{q}^* \right\|_2 ,$$

which indicates that $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is also an optimal solution of Problem (2.4). Meanwhile, note that $\operatorname{rank}(\mathbf{Q}') = \operatorname{rank}(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) \le \min\{\operatorname{rank}(\mathbf{V}), \operatorname{rank}(\mathbf{Y}_{11})\} \le K$, $\boldsymbol{\delta} = \boldsymbol{\kappa}_0$, and $\boldsymbol{\nu}_k = \boldsymbol{\kappa}_k$ for any $k \in [K]$. Thus, the proof is complete. □

When $K \ge m$, we have $\operatorname{rank}(\mathbf{Q}') \le m \le K$, thereby no need to consider this case in Theorem 2. Note that K is the number of pieces formulating the piecewise linear function $f(\mathbf{x}, \boldsymbol{\xi})$ and it is usually small for practical problems. For instance, in the CVaR and newsvendor problems, we have $K = 2$ (see Example 1 and Section 2.7.1.1, respectively). Thus, Theorem 2 shows that the rank of $\mathbf{Q}'$ that optimizes Problem (2.4), i.e., K, is usually small. We then expect that for any $m_1 \in [m]$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, the rank of the optimal $\mathbf{Q}_\mathrm{r}$ in Problem (2.13) might also be no greater than K and hence small for practical problems. With Problem (2.12), we then would like to choose a small $m_1 \ge K$ and find a $\mathbf{B} \in \mathcal{B}_{m_1}$ such that $\Theta_\mathrm{L}(m_1)$ can be close to $\Theta_\mathrm{M}(m)$.

We used to conjecture that the optimal value of Problem (2.12) equals that of Problem (2.4) when $m_1 \ge K$. Most numerical experiments (see Section 2.7) show this conjecture may be correct, while we also find a feasible solution of Problems (2.12) and (2.13) such that the corresponding objective value is equal to the optimal value of the original Problem (2.4) (see Theorem 9 in Appendix A.3.4). Nevertheless, we find an example to illustrate that the optimal value of Problem (2.12) with $m_1 = K$ can be strictly less than the optimal value of Problem (2.4) (see Example 2 in Appendix A.3.5). Thus, while the optimized dimensionality reduction maintains very high-quality solutions (mostly the optimal solutions as shown in our later numerical experiments in Section 2.7), we may still potentially lose some useful information that achieves the optimal solution of the original problem. To resolve this issue, we will also derive an upper bound and a new lower bound for the optimal value of the original problem in the later sections.

Note that Problem (2.12) is a nonconvex optimization problem due to the max-min operator. That is, we develop a low-dimensional nonconvex optimization technique to solve the original high-dimensional SDP problem, which can be significantly difficult to solve because of the large sizes of SDP matrices. To further efficiently solve Problem (2.12), we first reformulate it into a bilinear SDP problem (see Proposition 2) under the following assumption and then propose efficient algorithms (see Section 2.6) to solve it.

Assumption 2. *The set $\mathcal{X}$ is convex with at least one interior point. More specifically, we consider the convex set $\mathcal{X}$ in a generic SDP form: $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n (\boldsymbol{\Delta}_i x_i) + \boldsymbol{\Delta}_0 \succeq 0\}$, where $\boldsymbol{\Delta}_i \in \mathbb{R}^{\tau\times\tau}$ is symmetric for any $i \in \{0, 1, \ldots, n\}$ and some $\tau \geq 1$.*

We use $\mathbf{a}_{ij}\mathbf{x} + a_{ij}^0$ ($\forall i \in [\tau], j \in [\tau]$) to denote the elements of the matrix $\sum_{i=1}^n (\boldsymbol{\Delta}_i x_i) + \boldsymbol{\Delta}_0$, where $\mathbf{a}_{ij}^\top \in \mathbb{R}^n$. We let $y_k^0(\mathbf{x}) = \mathbf{w}_k^0\mathbf{x} + d_k^0$ and $y_k(\mathbf{x}) = (\mathbf{w}_k^1\mathbf{x} + d_k^1, \ldots, \mathbf{w}_k^m\mathbf{x} + d_k^m)^\top = \mathbf{W}_k\mathbf{x} + \mathbf{d}_k$ for any $k \in [K]$, where $(\mathbf{w}_k^i)^\top \in \mathbb{R}^n$ for any $i \in \{0, 1, \ldots, m\}$ and $k \in [K]$, $\mathbf{W}_k \in \mathbb{R}^{m\times n}$ for any $k \in [K]$, and $\mathbf{d}_k \in \mathbb{R}^m$ for any $k \in [K]$. The following proposition holds.

Proposition 2. *Under Assumption 2, Problem (2.12) has the same optimal value as the following bilinear SDP formulation:*

$$\Theta_{\mathrm{L}}(m_1) = \max_{\substack{t_k, \mathbf{p}_k, \\ \mathbf{P}_k, \forall k\in[K], \\ \mathbf{Z}, \mathbf{B}}} \sum_{k=1}^K \left(t_k d_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \right) \mathbf{d}_k \right) - \sum_{i=1}^{\tau}\sum_{j=1}^{\tau} z_{ij} a_{ij}^0 \tag{2.25a}$$

$$\text{s.t.} \quad 1 - \sum_{k=1}^K t_k = 0, \ \sqrt{\gamma_1} - \left\| \sum_{k=1}^K \mathbf{p}_k \right\|_2 \geq 0, \ \gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^K \mathbf{P}_k = 0, \tag{2.25b}$$

$$t_k \left(\mathbf{A}\boldsymbol{\mu} - \mathbf{b} \right)^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B} \right)^\top \mathbf{A}^\top \leq 0, \ \forall k \in [K], \tag{2.25c}$$

$$\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}=0, \tag{2.25d}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{Z}\succeq 0, \tag{2.25e}$$

where $\mathbf{p}_k\in\mathbb{R}^{m_1}$ *(*$k\in[K]$*),* $\mathbf{P}_k\in\mathbb{R}^{m_1\times m_1}$ *(*$k\in[K]$*),* $\mathbf{Z}\in\mathbb{R}^{\tau\times\tau}$*,* $\mathbf{B}\in\mathbb{R}^{m\times m_1}$*, and* z_{ij} *is the element of the matrix* $\mathbf{Z}$*. In addition,* $\mathbf{Z}$ *is the dual variable of the constraint* $\sum_{i=1}^{n}(\boldsymbol{\Delta}_i x_i)+\boldsymbol{\Delta}_0\succeq 0$ *in* $\mathcal{X}$ *and* $\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix}$ *(*$\forall k\in[K]$*) are the dual variables of constraints* (2.13b)*.*

Although solving Problem (2.25) may not achieve the optimal value of Problem (2.4), Theorem 9 demonstrates that we are not far from our target to close the approximation gap and motivates us to further develop an upper bound for the optimal value of Problem (2.4) while *closing the gap* between them in the next section.

2.4 Upper Bound

Motivated by the benefits of introducing a decision variable $\mathbf{B}\in\mathcal{B}_{m_1}$ in (2.9) to develop an outer approximation, we develop an inner approximation for Problem (2.3) in this section by relaxing the second-moment constraint in $\mathcal{D}_{\mathrm{M}}$ while optimizing the choice of components in $\boldsymbol{\xi}_{\mathrm{I}}$ to be relaxed, leading to the best possible upper bound for the optimal value of Problem (2.3). Specifically, given $m_1\in[m]$, we build the following optimized inner approximation of Problem (2.3):

$$\Theta_{\mathrm{U}}(m_1):=\min_{\mathbf{B}\in\mathcal{B}_{m_1}}\ \overline{\Theta}(m_1,\mathbf{B}), \tag{2.26}$$

where

$$\overline{\Theta}(m_1,\mathbf{B})=\min_{\mathbf{x}\in\mathcal{X}}\ \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{U}}}\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x},\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)\right] \tag{2.27}$$

with

$$\mathcal{D}_{\mathrm{U}}(\mathcal{S}_{\mathrm{I}},\gamma_1,\gamma_2)=\left\{\mathbb{P}_{\mathrm{I}}\;\middle|\;\begin{array}{l}\mathbb{P}_{\mathrm{I}}\left(\boldsymbol{\xi}_{\mathrm{I}}\in\mathcal{S}_{\mathrm{I}}\right)=1,\quad \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^{\top}\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\Big[\boldsymbol{\xi}_{\mathrm{I}}\Big]\leq\gamma_1\\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}^{\top}\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}\mathbf{B}\right]\preceq\gamma_2\mathbf{I}_{m_1}\end{array}\right\}. \tag{2.28}$$

The second-moment constraint in $\mathcal{D}_{\mathrm{U}}$ is relaxed from $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}]\preceq\gamma_2\mathbf{I}_m$, where we introduce a decision variable $\mathbf{B}\in\mathcal{B}_{m_1}$ to *optimize such a relaxation* of the ambiguity set, leading to optimized dimensionality reduction over this second-moment constraint. Intuitively, the feasible region defined by the ambiguity set $\mathcal{D}_{\mathrm{U}}$ is larger than that by $\mathcal{D}_{\mathrm{M}}$. Therefore, we have several conclusions based on this new ambiguity set $\mathcal{D}_{\mathrm{U}}$.

Theorem 3. *The following three conclusions hold: (i) Problem* (2.26) *provides an upper bound for the optimal value of Problem* (2.3)*, i.e.,* $\Theta_{\mathrm{U}}(m_1)\geq\Theta_{\mathrm{M}}(m)$ *for any* $m_1\leq m$*; (ii) the optimal value of Problem* (2.26) *is nonincreasing in* m_1*, i.e.,* $\Theta_{\mathrm{U}}(m_1)\geq\Theta_{\mathrm{U}}(m_2)$ *for any* $m_1<m_2\leq m$*; and (iii) when* $m_1=m$*, Problem* (2.26) *and Problem* (2.3) *have the same optimal value, i.e.,* $\Theta_{\mathrm{U}}(m)=\Theta_{\mathrm{M}}(m)$*.*

Theorem 3 shows that Problem (2.26) provides a valid upper bound for the optimal value of Problem (2.3), $\Theta_{\mathrm{M}}(m)$, and the upper bound is closer to $\Theta_{\mathrm{M}}(m)$ if less information is relaxed in $\mathcal{D}_{\mathrm{U}}$.

Proposition 3. *Under Assumption 1, Problem* (2.27) *has the same optimal value as the following SDP formulation:*

$$\begin{aligned}
\overline{\Theta}(m_1,\mathbf{B})=\quad\min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}},\hat{\mathbf{u}}}}\quad & \phi(m_1,s,\mathbf{q},\mathbf{Q}_{\mathrm{r}}) && \text{(2.29a)}\\
\text{s.t.}\quad & \begin{bmatrix}\chi(k,\mathbf{x},s,\boldsymbol{\lambda}_k) & \frac{1}{2}\mathbf{u}_k^{\top}\\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}}\end{bmatrix}\succeq 0,\ \forall k\in[K], && \text{(2.29b)}\\
& \mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\top}\boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k)=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K], && \text{(2.29c)}\\
& \mathbf{x}\in\mathcal{X},\ \mathbf{q}\in\mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}}\in\mathbb{R}^{m_1\times m_1}, && \text{(2.29d)}\\
& \hat{\boldsymbol{\lambda}}=\{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_K\},\boldsymbol{\lambda}_k\in\mathbb{R}_+^l,\hat{\mathbf{u}}=\{\mathbf{u}_1,\ldots,\mathbf{u}_K\},\mathbf{u}_k\in\mathbb{R}^{m_1},\forall k\in[K].
\end{aligned}$$

$$(2.29e)$$

Proposition 3 shows that Problem (2.26) can be updated by replacing its inner optimization Problem (2.27) with Problem (2.29). More importantly, Problem (2.26) becomes a low-dimensional SDP formulation, which can be solved more efficiently than solving the original high-dimensional formulation (2.4). With the updated Problem (2.26), we have the following conclusion.

Theorem 4. *Consider the optimal solution* $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}', \mathbf{Q}')$ *of Problem* (2.4)*,* S_k $(\forall k \in [K])$*,* $\mathbf{V}$*,* $\boldsymbol{\delta}$*,* $\mathbf{Y}_{11}$*, and* $\boldsymbol{\nu}_k$ $(\forall k \in [K])$ *that are defined in Theorem 2. If* $m_1 \geq K$*, then* $\Theta_{\mathrm{U}}(m_1) = \Theta_{\mathrm{M}}(m)$*.*

Theorem 4 shows that when $m_1 \geq K$, the optimal value of Problem (2.26) is always equal to the optimal value of the original problem, $\Theta_{\mathrm{M}}(m)$. Specifically, when $m_1 = K$, there exist optimal $\mathbf{B} = \mathbf{V}$ and $\mathbf{Q}_{\mathrm{r}} = \mathbf{Y}_{11}$ in Problem (2.26) such that $\Theta_{\mathrm{U}}(m_1) = \overline{\Theta}(m_1, \mathbf{V})$. Recall that K represents the number of pieces in the piecewise linear objective function $f(\mathbf{x}, \boldsymbol{\xi})$ and also determines the rank of the optimal $\mathbf{Q}$ in the original Problem (2.4) (see Theorem 2). Thus, we can decompose an optimal $\mathbf{Q}$ with a rank of K into $\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, where $\mathbf{V} \in \mathbb{R}^{m \times K}$ such that $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$ and $\mathbf{Y}_{11} \in \mathbb{R}^{K \times K}$. In Problem (2.26), when $m_1 = K$, we have $\mathbf{B} \in \mathbb{R}^{m \times K}$ and $\mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{K \times K}$ in the subproblem (2.29), by which constructing a solution with $\mathbf{B} = \mathbf{V}$ and $\mathbf{Q}_{\mathrm{r}} = \mathbf{Y}_{11}$ shows that $\Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m)$. Clearly, such a tightness result also holds when $m_1 > K$ by the monotonicity result in Theorem 3.

We may further interpret the insights as follows. Comparing the inner-approximation Problem (2.26) and the original Problem (2.3), we can notice that they differ only in the second-moment constraints in their ambiguity sets. When $m_1 = K$, we relax the original second-moment constraint from $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2\mathbf{I}_m$ to $\mathbf{B}^\top\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbf{B} \preceq \gamma_2\mathbf{I}_K$ with $\mathbf{B} \in \mathcal{B}_K$. That is, when such a relaxation is jointly optimized via the decision $\mathbf{B}$ with the subsequent SDP reformulation, it eventually does not lead to a different optimal value.

Specifically, under a worst-case distribution $\mathbb{P}_{\mathrm{I}}^*$ generated by solving Problem (2.3), we have $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$ may be equivalent to $\mathbf{B}^\top \mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbf{B} \preceq \gamma_2 \mathbf{I}_K$. Such equivalence largely depends on the low-rank property provided by Theorem 2, which states that the rank of an optimal solution of $\mathbf{Q}$ of Problem (2.4) is not larger than K. Note that the variable $\mathbf{Q}$ in Problem (2.4) is a dual variable with respect to the second-moment constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$, indicating that the rank of $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]$ may not be large. More specifically, we have the following proposition holds.

Proposition 4. *For any PSD matrix $\mathbf{X} \in \mathbb{R}^{m\times m}$ such that $\mathrm{rank}(\mathbf{X}) \le m_1 \le m$, we have the following equivalence holds:*

$$\mathbf{X} \preceq \mathbf{I}_m \iff \left(\mathbf{B}^\top \mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1},\ \forall \mathbf{B} \in \mathcal{B}_{m_1}\right).$$

Corollary 1. *For any PSD matrix $\mathbf{X} \in \mathbb{R}^{m\times m}$ and $\mathrm{rank}(\mathbf{X}) \le m_1 \le m$, there exists a $\mathbf{B} \in \mathcal{B}_{m_1}$ such that $\mathbf{X} \preceq \mathbf{I}_m$ is equivalent to $\mathbf{B}^\top \mathbf{X}\mathbf{B} \preceq \mathbf{I}_{m_1}$.*

In the context of solving Problem (2.3) and its inner-approximation Problem (2.26), Proposition 4 and Corollary 1 show that there exist a worst-case distribution $\mathbb{P}_{\mathrm{I}}^* \in \mathcal{D}_{\mathrm{M}}$ such that the rank of $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]$ is not larger than K and an optimal solution $\mathbf{B}^* \in \mathcal{B}_{m_1}$ such that $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top] \preceq \gamma_2 \mathbf{I}_m$ is equivalent to $(\mathbf{B}^*)^\top \mathbb{E}_{\mathbb{P}_{\mathrm{I}}^*}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top]\mathbf{B}^* \preceq \gamma_2 \mathbf{I}_K$. As such, even when we use a relaxed second-moment constraint, Problem (2.26) with $m_1 \ge K$ does not lose the optimality.

2.5 Lower Bound Revisited

Given that Problem (2.26) with $m_1 = K$ maintains the optimal value of the original Problem (2.3), we can further perform dimensionality reduction based on Problem (2.26) as we do in Section 2.3, thereby obtaining a new lower bound for the optimal value of Problem (2.3). Specifically, we consider $K \le m$ and recall that $\mathcal{B}_K = \{\mathbf{B} \in \mathbb{R}^{m\times K} \mid \mathbf{B}^\top \mathbf{B} = \mathbf{I}_K\}$.

Given any $m_1 \in [K]$, we consider

$$\min_{\mathbf{B}\in\mathcal{B}_K} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{L2}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \tag{2.30}$$

with

$$\mathcal{D}_{\mathrm{L2}}(\mathcal{S}_{\mathrm{I}}, \gamma_1, \gamma_2) = \left\{\mathbb{P}_{\mathrm{I}} \;\middle|\; \begin{array}{l} \mathbb{P}_{\mathrm{I}}\left(\boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}\right) = 1, \quad \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}^\top\right]\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\boldsymbol{\xi}_{\mathrm{I}}\right] \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1\right] \preceq \gamma_2\mathbf{I}_{m_1}, \quad \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2\right] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)} \end{array}\right\},$$

$\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2]$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, and $\mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)}$. To obtain the above ambiguity set $\mathcal{D}_{\mathrm{L2}}$, we shrink the ambiguity set $\mathcal{D}_{\mathrm{U}}$ of Problem (2.26) by replacing the second-moment constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}] \preceq \gamma_2\mathbf{I}_K$ with $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1] \preceq \gamma_2\mathbf{I}_{m_1}$ and $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)}$. The constraint $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2] \preceq \mathbf{0}_{(K-m_1)\times(K-m_1)}$ implies that we project the random vector $\boldsymbol{\xi}_{\mathrm{I}}$ to the space spanned by the columns of $\mathbf{B}_2$ and the second-moment value of the projected random vector is fixed at 0. By doing so, we may slightly lose some information to characterize the distribution $\mathbb{P}_{\mathrm{I}}$, but we can obtain a formulation with an even smaller size than Problem (2.26) and maintain high-quality solutions. Specifically, the following theorem holds.

Theorem 5. *Under Assumption 1, by dualizing the inner maximization problem of Problem (2.30), we obtain the following SDP formulation:*

$$\Theta_{\mathrm{L2}}(m_1) = \min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}}',\hat{\mathbf{u}}',\hat{\mathbf{u}}'',\\ \mathbf{B}_1,\mathbf{B}_2}} \quad \phi(m_1, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}) \tag{2.31a}$$

$$\text{s.t.} \quad \begin{bmatrix} \chi(k, \mathbf{x}, s, \boldsymbol{\lambda}_k) & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_{\mathrm{r}}' \end{bmatrix} \succeq 0, \;\; \forall k \in [K], \tag{2.31b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \boldsymbol{\psi}(k, \mathbf{x}, \boldsymbol{\lambda}_k) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'', \;\; \forall k \in [K], \tag{2.31c}$$

$$\mathbf{x} \in \mathcal{X}, \;\; \mathbf{q} \in \mathbb{R}^m, \;\; \mathbf{Q}_{\mathrm{r}}' \in \mathbb{R}^{m_1\times m_1}, \tag{2.31d}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)}, \quad [\mathbf{B}_1, \mathbf{B}_2]^\top[\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K,$$

$$(2.31e)$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}\,,\ \boldsymbol{\lambda}_k \in \mathbb{R}^l_+,\ \forall k \in [K], \tag{2.31f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}'_1, \ldots, \mathbf{u}'_K\}\,,\ \mathbf{u}'_k \in \mathbb{R}^{m_1},\ \forall k \in [K], \tag{2.31g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}''_1, \ldots, \mathbf{u}''_K\}\,,\ \mathbf{u}''_k \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. \tag{2.31h}$$

In addition, the following three conclusions hold: (i) Problem (2.31) provides a lower bound for the optimal value of Problem (2.4), i.e., $\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{M}}(m)$ for any $m_1 \leq K$; (ii) the optimal value of Problem (2.31) is nondecreasing in m_1, i.e., $\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{L2}}(m_2)$ for any $m_1 < m_2 \leq K$; and (iii) when $m_1 = K$, Problem (2.31) and Problem (2.4) have the same optimal value, i.e., $\Theta_{\mathrm{L2}}(K) = \Theta_{\mathrm{M}}(m)$.

Recall that the lower bound provided by Problem (2.12) may not achieve the optimal value of the original Problem (2.4) when reducing the dimensionality to K. However, the new lower bound provided by Problem (2.31) achieves the optimal value of the original problem when $m_1 = K$.

2.6 Efficient Algorithm

In Sections 2.3–2.5, we provide two outer approximations (i.e., Problems (2.25) and (2.31)) leading to lower bounds for the optimal value of Problem (2.3) and an inner approximation (i.e., Problem (2.26)) leading to an upper bound. Both Problems (2.26) and (2.31) can achieve the same optimal value as the original problem when $m_1 = K$. Note that the three approximations are low-dimensional bilinear SDP problems, which are nonconvex. It is hard to obtain the optimal solution of a bilinear SDP problem. To develop computationally efficient algorithms for bilinear SDPs, we derive Alternating Direction Method of Multipliers (ADMM) algorithms (Hajinezhad and Shi 2018, Wang et al. 2019, Themelis and Patrinos 2020) to solve the three approximations (see Appendix A.6.1 for more detailed reasons for choosing ADMM rather than other techniques). Meanwhile, it is important to note that our proposed algorithms are used to obtain a near-optimal dimensionality reduction matrix $\mathbf{B}$. Other solutions (e.g., $\mathbf{x}$) returned by the ADMM algorithm are not used here. Given this near-optimal $\mathbf{B}$, we can solve the low-dimensional SDP problems to retrieve the efficient lower and upper bounds for the original optimal value (see details in Section 2.6.2).

Note that Problems (2.25), (2.26), and (2.31) share the following generic formulation:

$$
\begin{aligned}
\min_{\mathbf{B},\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]} \quad & g\left(\mathbf{x},\mathbf{u}_k,\tilde{\mathbf{u}}_k,\forall k\in[K]\right) && \text{(2.32a)}\\
\text{s.t.} \quad & \left(\mathbf{x},\mathbf{u}_k,\tilde{\mathbf{u}}_k,\ \forall k\in[K]\right)\in\mathcal{U}, && \text{(2.32b)}\\
& \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}, && \text{(2.32c)}\\
& \tilde{\mathbf{u}}_k=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K], && \text{(2.32d)}
\end{aligned}
$$

where $\mathbf{B}\in\mathbb{R}^{m\times m_1}$, $\mathbf{u}_k\in\mathbb{R}^{m_1}$, $\tilde{\mathbf{u}}_k\in\mathbb{R}^{m}$, $\forall k\in[K]$, $\mathcal{U}$ is a convex set with at least one interior point, and $g(\cdot)$ is a differentiable convex function. Note that constraints (2.32c) and (2.32d) are bilinear constraints. We consider the following augmented Lagrangian problem for Problem (2.32):

$$
\min_{\substack{\mathbf{x},\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K];\\ \mathbf{B};\ \boldsymbol{\beta}_k,\forall k\in[K]}} \left\{ g\left(\mathbf{x},\mathbf{u}_k,\tilde{\mathbf{u}}_k,\forall k\in[K]\right)+\sum_{k=1}^{K}\boldsymbol{\beta}_k^\top\left(\tilde{\mathbf{u}}_k-\mathbf{B}\mathbf{u}_k\right) + \sum_{k=1}^{K}\frac{\rho_k}{2}\left\|\tilde{\mathbf{u}}_k-\mathbf{B}\mathbf{u}_k\right\|_2^2 \;\middle|\; \text{(2.32b)}-\text{(2.32c)} \right\}, \tag{2.33}
$$

where $\boldsymbol{\beta}_k\in\mathbb{R}^m$ ($\forall k\in[K]$) are Lagrangian multipliers and $\rho_k>0$ ($\forall k\in[K]$) are the penalty parameters. Thus, we design Algorithm 1 to solve Problem (2.33). Thereafter, we use the superscript i to denote the iteration step of Algorithm 1.

In Algorithm 1, we initialize $\mathbf{B}^0=\begin{bmatrix}\mathbf{I}_{m_1}\\ \mathbf{0}_{(m-m_1)\times m_1}\end{bmatrix}$ based on the PCA approximation in Cheramin et al. (2022) and set $\boldsymbol{\beta}_k^0=\mathbf{0}$ for any $k\in[K]$. We terminate the iteration when the improvement (regarding the relative gap) of the objective value is less than 10^{-4}. In this algorithm, given $\mathbf{B}$ and $\boldsymbol{\beta}_k$ for any $k\in[K]$, Problem (2.33) becomes a low-

Algorithm 1 ADMM for Problem (2.32)

Initialize: $\mathbf{B}^0, \boldsymbol{\beta}_k^0, \forall k \in [K]$

Repeat: update $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \mathbf{B}$ and $\boldsymbol{\beta}_k(\forall k \in [K])$ alternatingly by

Given $\mathbf{B}^i$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (2.33) to obtain the optimal solution $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$;

Given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (2.33) to obtain the optimal solution $\mathbf{B}^{i+1}$;

$\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho_k^i \left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}\right), \ \forall k \in [K];$

Until Convergence.

dimensional (i.e., $m_1 + 1$) SDP problem. Given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, Problem (2.33) becomes a nonconvex optimization problem, while the following proposition shows that it has an analytical optimal solution. Thus, Algorithm 1 can be performed efficiently.

Proposition 5. *Given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, Problem (2.33) has an optimal solution $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$, where $\sum_{k=1}^K (\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho_k \tilde{\mathbf{u}}_k \mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$ for $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times m_1}$, $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{m_1 \times m_1}$, and $\tilde{\mathbf{V}} \in \mathbb{R}^{m_1 \times m_1}$ by the singular value decomposition (SVD).*

2.6.1 Convergence Analyses

We further analyze the convergence property of Algorithm 1 to ensure the dimensionality reduction solution $\mathbf{B}$ returned by this algorithm is near-optimal, i.e., a theoretical guarantee. First, the following lemma holds.

Lemma 3. *Given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, we have $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ (an optimal solution of Problem (2.33)) is also an optimal solution of the following convex optimization problem:*

$$\max_{\mathbf{B} \in \mathbb{R}^{m \times m_1}} \left\{ \sum_{k=1}^K \left(\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top\right) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top \mathbf{B} \preceq \mathbf{I}_{m_1} \right\}. \tag{2.34}$$

To the best of our knowledge, no proof of convergence exists for the proposed ADMM algorithm in our specific setting. Existing convergence results focus on different problem settings. For instance, Hajinezhad and Shi (2018) employ the ADMM algorithm to solve a general bilinear optimization problem. Similar to our approach, they penalize the linking constraint and incorporate other constraints using an indicator function. However, in their setting, all the constraints except the linking constraint are convex, whereas we need to further deal with the nonconvex constraint $\mathbf{B}^\top \mathbf{B} = \mathbf{I}$. In addition, Chen and Goulart (2023) design an ADMM algorithm for diagonally constrained SDPs. Due to such a specific structure, they can derive the analytical optimal solution at each step. In contrast, the first step of our ADMM algorithm involves solving a general SDP problem for which no analytical optimal solution can be obtained.

Next, we present the convergence properties of our proposed ADMM algorithm. We let

$$\begin{aligned}
&\mathcal{L}\left(\mathbf{B}, \left(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right), \left(\boldsymbol{\beta}_k, \forall k \in [K]\right)\right) = \\
&\qquad g\left(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]\right) + \sum_{k=1}^{K} \boldsymbol{\beta}_k^\top \left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right) + \sum_{k=1}^{K} \frac{\rho_k}{2} \left\|\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right\|_2^2,
\end{aligned}$$

and introduce the following assumption commonly used in the convergence analysis of ADMM (Luo et al. 2008, Shen et al. 2014, Bertsekas 2014, Bai et al. 2021).

Assumption 3. *Given any $k \in [K]$, the sequence $\{\boldsymbol{\beta}_k^i\}$ is bounded and $\sum_{i=0}^{\infty} \|\boldsymbol{\beta}_k^{i+1} - \boldsymbol{\beta}_k^i\|_2^2 < \infty$.*

Because every bounded sequence has a convergent subsequence, we have

$$\boldsymbol{\beta}_k^i \to \boldsymbol{\beta}_k^*, \;\; i \to \infty, i \in \mathcal{I}, \;\; \forall k \in [K], \tag{2.35}$$

where $\{\beta_k^i\}_{i \in \mathcal{I}}$ is a subsequence of $\{\beta_k^i\}$. Meanwhile, under Assumption 3, by the update rule $\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1})$, we have

$$\boldsymbol{\beta}_k^{i+1} - \boldsymbol{\beta}_k^i = \rho_k^i(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}) \to 0,\ i \to \infty,\ \forall k \in [K].$$

Because $\rho_k^i > 0$ for any $k \in [K]$, we have

$$\tilde{\mathbf{u}}_k^i - \mathbf{B}^i\mathbf{u}_k^i \to 0,\ i \to \infty,\ \forall k \in [K]. \tag{2.36}$$

We are now ready to state the convergence theorem of Algorithm 1 as follows.

Theorem 6. *Let* $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *be any accumulation point of the sequence* $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ *generated by Algorithm 1. Then,* $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *satisfies the first-order stationary conditions of Problem* (2.32).

Note that one sufficient condition to ensure the existence of the accumulation point is that, at each iteration step $i+1$, given $\mathbf{B}^i$ and $(\boldsymbol{\beta}_k^i, \forall k \in [K])$ obtained from the previous iteration i, the optimal solution of the following problem is bounded:

$$\min_{\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]} \left\{ \mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \;\middle|\; (2.32\text{b}) \right\}. \tag{2.37}$$

This implies that, at iteration step $i+1$, the optimal solution of Problem (2.37), i.e., $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$, is bounded. By Assumption 3, the sequence $\{\boldsymbol{\beta}_k^i\}$ is also bounded. It follows that $\mathbf{B}^{i+1}$ is also bounded because $\mathbf{B}^{i+1}$ depends on $\boldsymbol{\beta}_k^i$, $\mathbf{u}_k^{i+1}$, and $\tilde{\mathbf{u}}_k^{i+1}$ by Proposition 5. Therefore, the bounded sequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ must have accumulation points.

2.6.2 Quality of the Dimensionality Reduction Solution B

Although the optimal value of our proposed three approximations can provide efficient lower or upper bounds for the original optimal value, our derived ADMM algorithms may not produce an optimal solution for each corresponding approximation problem because all three approximations are nonconvex problems. Therefore, it is important to recover the efficient lower or upper bounds using the dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithms. Clearly, a better dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ leads to a better lower or upper bound. Thus, we focus on the near-optimal dimensionality reduction solution $\mathbf{B}^{\text{ADMM}}$ that solving these outer and inner approximations can produce, rather than the solution of $\mathbf{x}$ that the ADMM algorithm returns. Specifically, we use this $\mathbf{B}^{\text{ADMM}}$ to recover the lower or upper bound for the optimal value of Problem (2.3) as follows.

- For Problem (2.25) (the first outer approximation), given the $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (2.13) to retrieve the lower bound.
- For Problem (2.26) (the inner approximation), given the $\mathbf{B}^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (2.29) to retrieve the upper bound.
- For Problem (2.31) (the second outer approximation), given the $[\mathbf{B}_1, \mathbf{B}_2]^{\text{ADMM}}$ returned by the ADMM algorithm, we solve the low-dimensional SDP problem (2.13) to retrieve the lower bound.

Note that for Problem (2.31), given the $[\mathbf{B}_1, \mathbf{B}_2]^{\text{ADMM}}$ returned by the ADMM algorithm, solving the subproblem of Problem (2.31) may not provide a lower bound for the original optimal value. Therefore, to recover a lower bound, we solve the low-dimensional SDP problem (2.13) to retrieve the lower bound.

In addition, we can measure the quality of the $\mathbf{B}^{\text{ADMM}}$ by deriving a theoretical interval in which the optimal value of Problem (2.3) is located via the following proposition.

Proposition 6. *Given any $m_1 \in [m]$ and $\mathbf{B}' \in \mathcal{B}_{m_1}$, we use $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_{\text{r}}^*, \mathbf{Q}_{\text{r}}^*)$ to denote an optimal solution of Problem (2.13). Let $P = \sum_{k=1}^{K} (\gamma_2/4)\mathbf{I}_m \bullet \mathbf{M}_k$ and $S =$*

$\min\{S_k,\ \forall k \in [K]\}$, *where*

$$\mathbf{M}_k = \left(\mathbf{B}'\mathbf{q}_\mathrm{r}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)\left(\mathbf{B}'\mathbf{q}_\mathrm{r}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top,$$
$$S_k = s^* - y_k^0(\mathbf{x}^*) - \boldsymbol{\lambda}_k^{*\top}\mathbf{b} - y_k(\mathbf{x}^*)^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^{*\top}\mathbf{A}\boldsymbol{\mu},\ \forall k \in [K].$$

We have

$$\Theta_\mathrm{M}(m) - \underline{\Theta}(m_1, \mathbf{B}') \le \frac{P}{S}\mathbb{I}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{I}(\sqrt{P} - S \ge 0).$$

Given any feasible $\mathbf{B}' \in \mathcal{B}_{m_1}$, solving the low-dimensional SDP problem (2.13) to obtain $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_\mathrm{r}^*, \mathbf{Q}_\mathrm{r}^*)$ is efficient. Thus, by solving an easy problem, we can obtain a theoretical lower bound $\underline{\Theta}(m_1, \mathbf{B}')$ and upper bound $\underline{\Theta}(m_1, \mathbf{B}') + (P/S)\mathbb{I}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{I}(\sqrt{P} - S \ge 0)$ for the original Problem (2.3) and quantify the gap between them in Proposition 6. More specifically, we can obtain their theoretical relative gap (denoted by "Theoretical Gap") as follows:

$$\text{Theoretical Gap} = \frac{\frac{P}{S}\mathbb{I}(\sqrt{P} - S < 0) + (2\sqrt{P} - S)\mathbb{I}(\sqrt{P} - S \ge 0)}{|\underline{\Theta}(m_1, \mathbf{B}')|} \times 100\%. \tag{2.38}$$

Note that this theoretical gap can be applied to all the three proposed approximations (i.e., Problems (2.25), (2.26), and (2.31)) because the theoretical gap can be calculated for any $\mathbf{B}' \in \mathcal{B}_{m_1}$ and all three approximations can yield a high-quality dimensionality reduction matrix $\mathbf{B}^\mathrm{ADMM}$ satisfying $(\mathbf{B}^\mathrm{ADMM})^\top\mathbf{B}^\mathrm{ADMM} = \mathbf{I}_{m_1}$ by using our proposed ADMM algorithm.

Proposition 6 implies that a better $\mathbf{B}^\mathrm{ADMM}$ returned by our proposed ADMM algorithms leads to a tighter theoretical gap. Note that our proposed three approximations provide efficient lower and upper bounds for the original Problem (2.3) and two of them are exact when we reduce the dimensionality of the uncertain space from m to $m_1 = K$ (see Theorems 4 and 5). Thus, we expect to have a very high-quality dimensionality reduction solution $\mathbf{B}^{\mathrm{ADMM}}$ returned by our ADMM algorithms, thereby implying a high-quality theoretical gap (2.38). Our numerical results in Section 2.7.2 show that such a theoretical gap is usually less than 5%.

2.7 Numerical Experiments

We perform extensive numerical experiments to demonstrate the effectiveness of our ODR approach in solving two moment-based DRO problems: multiproduct newsvendor and production-transportation problems. The mathematical models are implemented in MATLAB R2022a (ver. 9.12) by the modeling language CVX (ver. 2.2) and solved by the Mosek solver (ver. 9.3.20) on a PC with 64-bit Windows Operating System, an Intel(R) Xeon(R) W-2195 CPU @ 2.30GHz processor, and a 128 GB of memory. The time limit for each run is set at two hours. In Section 2.7.1, we specify the proposed inner and outer approximations under the ODR approach in the context of the multiproduct newsvendor and production-transportation problems. In Section 2.7.2, we report and analyze all the numerical results. All the data instances and source code can be found in Jiang et al. (2025a).

2.7.1 Numerical Setup

2.7.1.1 Multiproduct Newsvendor Problem

In the deterministic multiproduct newsvendor problem (Cheramin et al. 2022), we consider m products and the demand for each product $i \in [m]$ is ξ_i. Given the wholesale, retail, and salvage prices: $\mathbf{c} \in \mathbb{R}_+^m$, $\mathbf{v} \in \mathbb{R}_+^m$, and $\mathbf{g} \in \mathbb{R}_+^m$, respectively, we decide an ordering amount $\mathbf{x} \in \mathbb{R}_+^m$ to minimize the total cost

$$f\left(\mathbf{x}, \boldsymbol{\xi}\right) = \max\left\{\left(\mathbf{c} - \mathbf{v}\right)^\top \mathbf{x}, \left(\mathbf{c} - \mathbf{g}\right)^\top \mathbf{x} + \left(\mathbf{g} - \mathbf{v}\right)^\top \boldsymbol{\xi}\right\}.$$

Note that this piecewise linear function $f(\mathbf{x}, \boldsymbol{\xi})$ has only two pieces, i.e., $K = 2$. When the demand $\boldsymbol{\xi}$ is uncertain and its probability distribution belongs to a distributional ambiguity set $\mathcal{D}_{\mathrm{M0}}$ as defined in Section 2.2, we obtain the following DRO counterpart:

$$\min_{\mathbf{x}\geq 0} \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}} \; \mathbb{E}_{\mathbb{P}}\left[\max\left\{(\mathbf{c}-\mathbf{v})^\top \mathbf{x}, (\mathbf{c}-\mathbf{g})^\top \mathbf{x} + (\mathbf{g}-\mathbf{v})^\top \boldsymbol{\xi}\right\}\right]. \tag{2.39}$$

We can apply Proposition 1 to reformulate Problem (2.39) and the proposed inner and outer approximations (i.e., Problems (2.25), (2.26), and (2.31)) to approximate it (see details in Appendix A.7.1).

When the dimension m of $\boldsymbol{\xi}$ is large, the original SDP reformulation of Problem (2.39) (i.e., Problem (A.64)) becomes very difficult to solve because of the large-scale SDP constraints. Nevertheless, our approximations under the ODR approach (i.e., Problems (A.65)–(A.67)) have SDP matrices with very small sizes (e.g., $K+1 = 3$), largely reducing the computational burden while maintaining the solution quality.

2.7.1.2 Production-Transportation Problem (Bertsimas et al. 2010)

The deterministic production-transportation problem considers G suppliers, each with normalized capacity one, and H customers, each with demand d_j ($\forall j \in [H]$) such that $\sum_{j=1}^{H} d_j \leq G$. Supplier $i \in [G]$ produces x_i goods with unit production cost c_i and delivers z_{ij} goods to customer $j \in [H]$ with unit transportation cost ξ_{ij}, thereby satisfying all customer demands and minimizing the total cost. We can formulate this problem as follows:

$$\min_{\mathbf{x},\mathbf{z}} \quad \mathbf{c}^\top \mathbf{x} + \boldsymbol{\xi}^\top \mathbf{z} \tag{2.40a}$$

$$\text{s.t.} \quad 0 \leq x_i \leq 1, \; \forall i \in [G], \tag{2.40b}$$

$$\sum_{i=1}^{G} z_{ij} = d_j, \forall j \in [H], \sum_{j=1}^{H} z_{ij} = x_i, \forall i \in [G], z_{ij} \geq 0, \forall i \in [G], j \in [H], \tag{2.40c}$$

where $\boldsymbol{\xi} = (\xi_{ij},\ \forall i \in [G],\ j \in [H])^\top$ and $\mathbf{z}_{ij} = (z_{ij},\ \forall i \in [G],\ j \in [H])^\top$. Following the same setting in Bertsimas et al. (2010), we consider $\boldsymbol{\xi}$ is random and its probability distribution $\mathbb{P}$ belongs to $\mathcal{D}_{\mathrm{M0}}$. Thus, we can derive the two-stage DRO counterpart:

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\mathbb{P} \in \mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}} \left[\mathcal{U} \left(\mathcal{Q} \left(\mathbf{x}, \boldsymbol{\xi} \right) \right) \right] \;\middle|\; (2.40\mathrm{b}) \right\}, \tag{2.41}$$

where $\mathcal{U}(\mathcal{Q}(\mathbf{x}, \boldsymbol{\xi})) = \max_{k \in [K]} \{\alpha_k \mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) + \beta_k\}$ is a convex nondecreasing disutility function that incorporates risk attitudes into the second-stage cost and $\mathcal{Q}(\mathbf{x}, \boldsymbol{\xi}) = \min_{\mathbf{z}} \{\boldsymbol{\xi}^\top \mathbf{z} \mid (2.40\mathrm{c})\}$. We can reformulate Problem (2.41) into an SDP problem and apply our inner and outer approximations to approximate it (see details in Appendix A.7.2).

When the dimension GH of $\boldsymbol{\xi}$ is large, the original SDP reformulation of Problem (2.41) (i.e., Problem (A.68)) becomes very difficult to solve due to the high-dimensional SDP constraints, while our approximations (i.e., Problems (A.70)–(A.72)) have SDP matrices with very small sizes (e.g., $K + 1$) and can be solved efficiently.

2.7.2 Numerical Results

We compare the performance of our ODR approach (that provides two lower bounds and one upper bound) with three benchmark approaches: (i) the Mosek solver with default settings, which can provide the optimal value of the original problem; (ii) the low-rank algorithm proposed by Burer and Monteiro (2003) to solve the SDP reformulation of the original problem, i.e., Problem (2.4), generating a lower bound for the optimal value of Problem (2.4); and (iii) the existing PCA approximation proposed by Cheramin et al. (2022), generating PCA-based lower and upper bounds. For the third benchmark, we consider the reduced dimension $m_1 \in \{100\% \times \dim(\boldsymbol{\xi}), 80\% \times \dim(\boldsymbol{\xi}), 60\% \times \dim(\boldsymbol{\xi}), 40\% \times \dim(\boldsymbol{\xi}), 20\% \times \dim(\boldsymbol{\xi}), K\}$, where $\dim(\boldsymbol{\xi})$ represents the dimension of the random vector $\boldsymbol{\xi}$. Note that $K = 2$ in the multiproduct newsvendor problem. For the production-transportation problem, we consider $K \in \{5, 10, 15\}$. Our proposed inner and outer approximations under the ODR approach are solved using Algorithm 1, providing near-optimal $\mathbf{B}^*$ and recovering valid lower and upper bounds (see Section 2.6.2). For all instances, through tuning a range of values, we set the initial Lagrangian multipliers ($\boldsymbol{\beta}_k, \forall k \in [K]$) to $\mathbf{0}$ and the penalty parameters ($\rho_k, \forall k \in [K]$) to 200 in Algorithm 1.

2.7.2.1 Instance Generation and Table Header Description

We consider various instances of the multiproduct newsvendor and production-transportation problems. In the multiproduct newsvendor problem, the mean and standard deviation of $\boldsymbol{\xi}$ are randomly generated from the intervals $[0, 10]$ and $[1, 2]$, respectively. We further generate a correlation matrix randomly using the MATLAB function "gallery('randcorr',n)" and then convert it to a covariance matrix. We follow Xu et al. (2018) to set the wholesale, retail, and savage prices as $c_i = 0.1(5+i-1)$, $v_i = 0.15(5+i-1)$, and $g_i = 0.05(5+i-1)$ for any $i \in [m]$, respectively. Meanwhile, we consider $m \in \{100, 200, 400, 800, 1200, 1600, 2000\}$ in this problem.

In the production-transportation problem, we follow Bertsimas et al. (2010) to randomly generate the locations of G suppliers and H customers from a unit square and use ξ_{ij} to measure the distance between supplier $i \in [G]$ and customer $j \in [H]$. For any $i \in [G]$ and $j \in [H]$, we generate $10,000$ samples of ξ_{ij} from the interval $[0.5\bar{\xi}_{ij}, 1.5\bar{\xi}_{ij}]$ and use them to estimate the mean, standard deviation, and covariance matrix of $\boldsymbol{\xi}$. Using γ_0 denoting the average transportation cost, we then generate production cost c_i and demand d_j uniformly on the intervals $[0.5\gamma_0, 1.5\gamma_0]$ and $[0.5G/H, G/H]$, respectively, for any $i \in [G]$ and $j \in [H]$. The disutility function $\mathcal{U}(x) = 0.25(e^{2x} - 1)$ and is approximated by a piecewise-linear function with K equidistant segments on the interval $[0, 1]$. Meanwhile, we consider $(G, H) \in \{(4, 25), (5, 20), (5, 40), (8, 25), (10, 40), (20, 30), (20, 40)\}$.

For any given value of m in the multiproduct newsvendor problem or (G, H) in the production-transportation problem, we randomly generate five instances and report the average performance in Tables 2.1–2.4. Here we describe several table headers that are shared by these tables. We use "Mosek" and "Low-rank" to represent the performance of the Mosek solver and the low-rank algorithm (Burer and Monteiro 2003), respectively. The abbreviations "LB," "UB," and "RLB" represent lower, upper, and revisited lower bounds, respectively. Specifically, the labels "ODR-LB," "ODR-UB," and "ODR-RLB" denote the lower-bound performance after solving the first outer approximation (2.25) with $m_1 = K$, the upper-bound performance after solving the inner approximation (2.26) with $m_1 = K$, and the other lower-bound performance after solving the second outer approximation (2.31) with $m_1 = K$, respectively. Note that using the dimensionality reduction solution $\mathbf{B}$ given by our proposed ADMM algorithms after solving each approximation, we retrieve the corresponding lower or upper bound following the recovery process described in Section 2.6.2. We use "PCA-100%," "PCA-80%," "PCA-60%," "PCA-40%," "PCA-20%," and "PCA-K" to denote the performance of the PCA approximation by Cheramin et al. (2022) when the reduced dimension m_1 equals $100\% \times \dim(\boldsymbol{\xi})$, $80\% \times \dim(\boldsymbol{\xi})$, $60\% \times \dim(\boldsymbol{\xi})$, $40\% \times \dim(\boldsymbol{\xi})$, $20\% \times \dim(\boldsymbol{\xi})$, and K, respectively. For instance, "PCA-20%" and "PCA-2" in Table 2.1 denote the performance of the case when $m_1 = 20\% \times \dim(\boldsymbol{\xi})$ and $m_1 = K = 2$, respectively.

In all the tables, we use "Size" to represent the value of m in the multiproduct newsvendor problem or (G, H) in the production-transportation problem and "Time" to represent the computational time in seconds required to solve each instance. Note that following the recovery process described in Section 2.6.2, the computational time includes two parts. The first part is the time spent using the ADMM algorithm to solve an approximation problem. The second part is the time spent solving the lower- or upper-bound subproblem. The lower-bound subproblem (2.13) and the upper-bound subproblem (2.29) are low-dimensional SDP problems. In all our computational experiments, the time spent on the second part is less than 1% of the time spent on the first part. Therefore, we report the total time without separately distinguishing between the two parts. We use "Gap1" (resp. "Gap2") to represent the relative gap in percentage between a lower (resp. an upper) bound and the optimal value provided by the Mosek solver. That is,

$$\text{Gap1} = \frac{\text{optimal value} - \text{lower bound}}{|\text{optimal value}|} \times 100\%, \;\; \text{Gap2} = \frac{\text{upper bound} - \text{optimal value}}{|\text{optimal value}|} \times 100\%.$$

We further use "Interval Gap" to represent the relative gap in percentage between a lower bound and an upper bound, i.e.,

$$\text{Interval Gap} = \frac{\text{upper bound} - \text{lower bound}}{|\text{upper bound}|} \times 100\%. \tag{2.42}$$

Specifically, for both the ODR approach and the low-rank algorithm, we take the objective value of "ODR-UB" as the value of "upper bound" in (2.42). For the PCA approximation approach, the value of "upper bound" in (2.42) is provided by this approach itself. The value of the "Theoretical Gap" for each instance is defined in (2.38).

To further illustrate how to calculate the "Interval Gap," we consider the "Interval Gap" in the row "ODR-LB" of Table 2.1 as an example. We first use our ADMM algorithm to solve the first outer approximation, i.e., Problem (2.25), and record the obtained dimensionality reduction solution $\mathbf{B}^{\text{LB}}$. Given this $\mathbf{B}^{\text{LB}}$, we solve Problem (2.13) and obtain $\underline{\Theta}(m_1, \mathbf{B}^{\text{LB}})$, which is a lower bound. We then use the ADMM algorithm to solve the inner approximation, i.e., Problem (2.26), and record $\mathbf{B}^{\text{UB}}$. Given this $\mathbf{B}^{\text{UB}}$, we solve Problem (2.29) and obtain $\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}})$, which is an upper bound. Thus, we calculate the "Interval Gap" by $(\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}}) - \underline{\Theta}(m_1, \mathbf{B}^{\text{LB}}))/|\overline{\Theta}(m_1, \mathbf{B}^{\text{UB}})| \times 100\%$.

Finally, we use "-" to represent that no result can be obtained within the time limit (i.e., two hours). For instance, the Mosek solver cannot solve the original problem to the optimality within two hours for the newsvendor problem with $m \geq 400$. Hence, we cannot obtain the value of "Gap1" for the "Mosek," "ODR-LB," and "ODR-RLB" approaches.

Table 2.1: Average Performance on the Newsvendor Problem

	Size (m)	100	200	400	800	1200	1600	2000
Mosek	Time (secs)	13.02	363.54	-	-	-	-	-
Low-rank	Gap1 (%)	2.52	1.79	-	-	-	-	-
	Time (secs)	0.26	0.80	5.46	47.34	110.33	309.00	825.62
	Interval Gap (%)	4.27	3.66	2.67	2.32	2.28	2.36	2.52
ODR-LB	Gap1 (%)	0.09	0.00	-	-	-	-	-
	Time (secs)	0.77	0.78	0.83	0.85	1.13	2.01	2.54
	Interval Gap (%)	1.81	1.83	1.44	1.44	1.56	1.73	1.96
	Theoretical Gap (%)	2.73	3.27	2.47	3.15	4.28	2.91	3.58
ODR-RLB	Gap1 (%)	0.03	0.03	-	-	-	-	-
	Time (secs)	1.95	2.60	4.33	9.75	20.83	38.36	56.68
	Interval Gap (%)	1.74	1.86	1.46	1.46	1.56	1.78	1.98
	Theoretical Gap (%)	1.92	2.36	1.27	2.44	1.90	3.01	2.58
ODR-UB	Gap2 (%)	1.68	1.80	-	-	-	-	-
	Time (secs)	1.95	2.60	4.33	9.75	20.83	38.36	56.68
	Theoretical Gap (%)	1.92	2.36	1.27	2.44	1.90	3.01	2.58
PCA-100%	Gap1 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	13.04	361.54	-	-	-	-	-
	Gap2 (%)	0.00	0.00	-	-	-	-	-
	Time (secs)	12.99	361.91	-	-	-	-	-
	Interval Gap (%)	0.00	0.00	-	-	-	-	-
PCA-80%	Gap1 (%)	0.50	0.31	-	-	-	-	-
	Time (secs)	5.05	120.72	3348.00	-	-	-	-
	Gap2 (%)	12.23	11.13	-	-	-	-	-
	Time (secs)	7.77	155.39	4793.72	-	-	-	-
	Interval Gap (%)	14.54	12.90	13.57	-	-	-	-
PCA-60%	Gap1 (%)	0.98	0.73	-	-	-	-	-
	Time (secs)	1.44	28.73	785.63	-	-	-	-
	Gap2 (%)	23.33	24.07	-	-	-	-	-
	Time (secs)	2.29	44.28	1196.98	-	-	-	-
	Interval Gap (%)	31.76	32.79	31.87	-	-	-	-
PCA-40%	Gap1 (%)	1.69	1.20	-	-	-	-	-
	Time (secs)	0.39	5.21	125.43	3351.00	-	-	-
	Gap2 (%)	35.79	35.94	-	-	-	-	-
	Time (secs)	0.57	8.03	177.96	5237.40	-	-	-
	Interval Gap (%)	58.50	58.26	56.45	57.18	-	-	-
PCA-20%	Gap1 (%)	2.71	1.90	-	-	-	-	-
	Time (secs)	0.15	0.43	6.49	136.97	971.60	3546.30	-
	Gap2 (%)	47.92	48.19	-	-	-	-	-
	Time (secs)	0.17	0.60	9.25	203.97	1340.46	4940.28	-
	Interval Gap (%)	97.74	84.05	90.65	93.17	92.38	94.27	-
PCA-2	Gap1 (%)	4.26	3.24	-	-	-	-	-
	Time (secs)	0.11	0.12	0.13	0.20	0.26	0.36	0.50
	Gap2 (%)	57.60	59.25	-	-	-	-	-
	Time (secs)	0.13	0.14	0.16	0.22	0.32	0.46	0.60
	Interval Gap (%)	147.12	154.40	141.39	149.18	146.92	150.96	153.57

Table 2.2: Average Performance on the Production-Transportation Problem with $K = 5$

	Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	56.04	70.10	1524.45	1612.68	-	-	-
Low-rank	Gap1 (%)	3.69	4.17	2.94	3.42	-	-	-
	Time (secs)	14.87	15.28	44.82	41.34	192.21	710.28	1678.21
	Interval Gap (%)	3.69	4.17	2.94	3.42	5.44	4.12	5.93
ODR-LB	Gap1 (%)	0.14	0.34	0.22	0.29	-	-	-
	Time (secs)	4.03	3.63	6.01	4.82	12.41	25.03	57.79
	Interval Gap (%)	0.15	0.34	0.43	0.29	0.52	0.55	0.51
	Theoretical Gap (%)	1.24	1.07	4.91	0.77	2.71	1.25	1.09
ODR-RLB	Gap1 (%)	0.01	0.02	0.01	0.00	-	-	-
	Time (secs)	5.42	5.36	22.40	20.86	123.64	330.29	665.43
	Interval Gap (%)	0.02	0.02	0.01	0.01	0.00	0.00	0.00
	Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
ODR-UB	Gap2 (%)	0.01	0.01	0.00	0.00	-	-	-
	Time (secs)	5.48	5.32	22.41	20.86	123.48	329.95	665.35
	Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	56.74	68.76	1521.32	1612.94	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	55.49	68.49	1521.04	1611.67	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.52	0.99	0.33	1.50	-	-	-
	Time (secs)	23.41	27.68	589.74	625.65	-	-	-
	Gap2 (%)	3.31	2.18	1.15	2.76	-	-	-
	Time (secs)	29.33	31.20	657.78	639.09	-	-	-
	Interval Gap (%)	3.63	3.08	1.44	4.12	-	-	-
PCA-60%	Gap1 (%)	1.60	2.61	1.26	3.41	-	-	-
	Time (secs)	8.54	8.41	143.60	115.54	3087.98	-	-
	Gap2 (%)	9.02	7.33	3.45	8.62	-	-	-
	Time (secs)	9.04	8.88	158.95	152.21	3484.45	-	-
	Interval Gap (%)	9.50	9.17	4.36	11.04	4.85	-	-
PCA-40%	Gap1 (%)	3.75	4.23	2.33	4.00	-	-	-
	Time (secs)	2.18	2.11	30.80	23.44	512.53	3004.06	-
	Gap2 (%)	14.05	11.58	6.57	9.96	-	-	-
	Time (secs)	2.61	2.37	37.08	30.60	633.00	3149.11	-
	Interval Gap (%)	15.42	14.04	7.94	12.66	8.52	5.05	-
PCA-20%	Gap1 (%)	4.45	5.41	3.59	4.16	-	-	-
	Time (secs)	0.74	0.69	6.21	5.59	38.03	205.25	741.71
	Gap2 (%)	17.07	12.66	9.25	10.08	-	-	-
	Time (secs)	1.11	1.01	7.31	6.26	64.54	276.63	883.55
	Interval Gap (%)	18.35	15.95	11.32	12.91	9.39	5.22	4.85
PCA-5	Gap1 (%)	5.35	5.73	4.55	4.39	-	-	-
	Time (secs)	0.38	0.36	0.52	0.45	0.66	0.79	1.00
	Gap2 (%)	17.99	15.35	12.71	10.08	-	-	-
	Time (secs)	0.78	0.72	2.35	2.08	10.41	27.98	68.39
	Interval Gap (%)	19.76	18.26	15.15	13.12	9.69	5.48	5.14

Table 2.3: Average Performance on the Production-Transportation Problem with $K = 10$

	Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	115.52	103.52	2866.54	2857.47	-	-	-
Low-rank	Gap1 (%)	3.41	3.92	3.21	4.08	-	-	-
	Time (secs)	17.67	18.41	56.22	54.17	227.71	929.00	1907.28
	Interval Gap (%)	3.41	3.92	3.21	4.08	5.27	6.2	4.62
ODR-LB	Gap1 (%)	0.12	0.19	0.13	0.25	-	-	-
	Time (secs)	6.65	6.41	11.11	10.31	31.42	61.41	131.10
	Interval Gap (%)	0.12	0.19	0.13	0.25	0.22	0.19	0.25
	Theoretical Gap (%)	1.78	1.97	2.96	1.19	2.10	1.49	1.22
ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	11.22	10.65	40.34	32.95	122.54	342.80	748.11
	Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	11.14	10.69	40.28	32.92	122.48	344.10	747.90
	Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	115.74	102.90	2852.43	2851.09	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	116.32	103.70	2853.10	2859.01	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.47	0.39	0.38	0.58	-	-	-
	Time (secs)	52.31	50.48	1101.61	979.90	-	-	-
	Gap2 (%)	0.04	0.04	0.02	0.02	-	-	-
	Time (secs)	58.12	58.88	1392.67	1364.82	-	-	-
	Interval Gap (%)	0.51	0.42	0.40	0.61	-	-	-
PCA-60%	Gap1 (%)	1.54	1.39	0.98	1.57	-	-	-
	Time (secs)	14.48	15.54	270.80	240.40	4182.33	-	-
	Gap2 (%)	0.96	0.34	0.34	0.40	-	-	-
	Time (secs)	19.64	20.83	455.00	440.12	4582.31	-	-
	Interval Gap (%)	2.46	1.72	1.32	1.95	2.61	-	-
PCA-40%	Gap1 (%)	2.30	2.31	1.76	2.18	-	-	-
	Time (secs)	4.19	4.04	53.67	45.22	831.61	4312.63	-
	Gap2 (%)	1.85	1.63	1.10	0.77	-	-	-
	Time (secs)	6.19	5.96	94.74	97.70	931.48	4792.10	-
	Interval Gap (%)	4.05	3.87	2.81	2.92	5.01	4.95	-
PCA-20%	Gap1 (%)	2.98	3.20	2.26	2.39	-	-	-
	Time (secs)	2.26	1.91	7.72	7.02	52.34	361.81	1038.72
	Gap2 (%)	2.24	2.44	1.71	1.07	-	-	-
	Time (secs)	3.83	3.52	18.70	18.07	71.38	428.93	1391.25
	Interval Gap (%)	5.09	5.49	3.90	3.42	6.21	5.48	4.29
PCA-10	Gap1 (%)	3.36	3.26	2.40	2.53	-	-	-
	Time (secs)	0.83	0.76	1.19	1.00	2.01	1.95	2.54
	Gap2 (%)	2.51	2.48	2.09	1.11	-	-	-
	Time (secs)	1.93	1.81	7.12	5.97	21.15	43.62	79.15
	Interval Gap (%)	5.72	5.59	4.40	3.60	6.36	7.28	6.04

Table 2.4: Average Performance on the Production-Transportation Problem with $K = 15$

	Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
Mosek	Time (secs)	154.22	159.12	4004.18	4099.14	-	-	-
Low-rank	Gap1 (%)	4.51	3.52	4.06	5.91	-	-	-
	Time (secs)	21.73	28.14	74.20	62.18	331.01	1124.41	2271.82
	Interval Gap (%)	4.51	3.52	4.06	5.91	4.22	5.11	4.83
ODR-LB	Gap1 (%)	0.05	0.10	0.12	0.15	-	-	-
	Time (secs)	13.92	16.56	22.11	26.16	43.02	80.51	168.90
	Interval Gap (%)	0.05	0.10	0.12	0.15	0.11	0.13	0.09
	Theoretical Gap (%)	4.69	5.22	5.57	6.38	3.29	4.41	3.95
ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	22.60	21.45	77.18	63.99	241.03	689.24	1550.21
	Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	22.61	21.44	76.99	64.08	149.21	401.63	878.68
	Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
PCA-100%	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	154.64	158.67	4024.62	4095.91	-	-	-
	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
	Time (secs)	156.32	155.64	4050.04	4059.01	-	-	-
	Interval Gap (%)	0.00	0.00	0.00	0.00	-	-	-
PCA-80%	Gap1 (%)	0.47	0.53	0.62	0.58	-	-	-
	Time (secs)	68.31	69.22	1439.48	1428.65	-	-	-
	Gap2 (%)	0.04	0.06	0.15	0.08	-	-	-
	Time (secs)	79.55	80.20	2120.31	2146.05	-	-	-
	Interval Gap (%)	0.51	0.57	0.75	0.62	-	-	-
PCA-60%	Gap1 (%)	0.95	1.13	1.28	1.25	-	-	-
	Time (secs)	22.50	25.11	411.27	398.64	-	-	-
	Gap2 (%)	0.22	0.31	0.47	0.62	-	-	-
	Time (secs)	33.28	38.11	658.16	624.21	-	-	-
	Interval Gap (%)	1.17	1.40	1.74	1.85	-	-	-
PCA-40%	Gap1 (%)	1.68	1.67	1.78	1.81	-	-	-
	Time (secs)	6.61	8.53	81.20	80.88	1553.12	-	-
	Gap2 (%)	1.65	2.58	2.54	1.96	-	-	-
	Time (secs)	10.21	12.37	137.98	155.84	2296.17	-	-
	Interval Gap (%)	3.23	4.27	4.31	3.76	4.53	-	-
PCA-20%	Gap1 (%)	2.22	3.01	2.09	1.91	-	-	-
	Time (secs)	4.74	4.98	13.21	14.20	81.28	375.81	1671.21
	Gap2 (%)	3.07	3.61	2.98	2.48	-	-	-
	Time (secs)	9.24	8.19	27.26	28.79	124.04	496.38	2517.39
	Interval Gap (%)	5.27	6.59	5.06	4.22	5.78	6.29	5.88
PCA-15	Gap1 (%)	2.24	3.32	2.94	3.09	-	-	-
	Time (secs)	1.91	2.07	2.52	2.73	4.25	4.87	6.27
	Gap2 (%)	3.10	4.12	3.57	4.01	-	-	-
	Time (secs)	4.74	5.12	12.86	12.92	23.16	48.21	95.72
	Interval Gap (%)	5.31	7.40	6.41	7.03	6.42	7.19	6.83

2.7.2.2 Numerical Performance

From Tables 2.1–2.4, we have the following observations. First, in the newsvendor problem with $m \in \{100, 200\}$ and the production-transportation problem with $(G, H) \in \{(4, 25), (5, 20), (5, 40), (8, 25)\}$, the Mosek solver solves each instance of the original problem to the optimality. Our ODR approach performs much better than the three benchmark approaches. Both the "ODR-LB" and "ODR-RLB" provide a smaller value of "Gap1" than the low-rank algorithm and the PCA approximation, and require a shorter computational time than the three benchmark approaches. The "ODR-UB" also provides a smaller value of "Gap2" than the PCA approximation if $m_1 \neq 100\% \times \dim(\boldsymbol{\xi})$ therein and requires shorter computational time.

Specifically, the objective value of our "ODR-LB" reaches the optimal value of the original problem for some instances of the multiproduct newsvendor problem (see Table 2.1) and provides near-optimal solutions for the production-transportation problem with "Gap1" less than 0.34% (see Tables 2.2–2.4). More importantly, the "ODR-LB" reduces the computational time by up to three orders of magnitude compared to the Mosek solver. In addition, the "ODR-RLB" and "ODR-UB" *reach the optimal value of the original problem for all instances* in Tables 2.3–2.4 and provide objective values that are near-optimal (within 1.8% for all instances and 0.03% for most instances) in Tables 2.1–2.2, while reducing the computational time significantly. The results imply that our ADMM algorithms return the optimal $\mathbf{B}^*$ for most instances and the numerical gap between our derived lower and upper bounds (i.e., "Interval Gap") can be tight.

In addition, Tables 2.1–2.4 show that our ODR approach (including "ODR-LB," "ODR-UB," and "ODR-RLB") provides a better solution in terms of the objective value than the PCA approximation if the reduced dimension $m_1 \leq 80\% \times \dim(\boldsymbol{\xi})$ in the latter approach. That is, even if we maintain 80% of the random parameters corresponding to the largest eigenvalues to be uncertain in the PCA approximation by focusing on only their statistical information, the performance is worse than our ODR approach, where we *optimize* the dimensionality reduction from $\dim(\boldsymbol{\xi})$ to K (i.e., maintaining only 1% of the original dimensionality size when $m = 200$ and $K = 2$ for the multiproduct newsvendor problem). More importantly, the inner and outer approximations of our ODR approach can be solved efficiently.

Second, when the problem size is large, i.e., $m \geq 400$ in the newsvendor problem and $(G, H) \in \{(10, 40), (20, 30), (20, 40)\}$ in the production-transportation problem, the Mosek solver cannot solve any instance of the original problem to the optimality. Our ODR approach also performs better than the three benchmark approaches. Tables 2.1–2.4 show that "ODR-LB" provides a smaller value of "Interval Gap" (within 2%) and requires a much shorter computational time than both the low-rank algorithm and the PCA approximation. For instance, when $m = 1600$ in the multiproduct newsvendor problem, the low-rank algorithm and "ODR-LB" take 309 and 2.01 seconds to solve an instance of the multiproduct newsvendor problem and provide the value of "Interval Gap" at 2.36% and 1.73%, respectively. The PCA approximation solves this instance only when the reduced dimension m_1 is not larger than $20\% \times m$, by which it takes 3546.3 seconds while the solution quality is very poor, providing the value of "Interval Gap" at 94.27%. Tables 2.2–2.4 show that the "ODR-RLB" and "ODR-UB" *reach the optimal value of the original problem for all instances* (with "Interval Gap" at 0); that is, our ADMM algorithms return the optimal $\mathbf{B}^*$. Meanwhile, the value of the "Theoretical Gap" is mostly within 5%. More importantly, our ODR approach is not sensitive to the value of $\dim(\boldsymbol{\xi})$, while the benchmark approaches perform much worse when $\dim(\boldsymbol{\xi})$ is larger. Thus, when we cannot obtain the optimal value of the original problem, the "ODR-LB", "ODR-RLB", and "ODR-UB" can be efficiently solved to provide a narrower interval that includes the optimal value than the benchmark approaches. That is, our ODR approach can provide a near-optimal solution very efficiently for the moment-based DRO problems where other best possible benchmark approaches are struggling.

Third, "ODR-RLB" leads to better solution quality than "ODR-LB". From Tables 2.1-

2.4, we observe that (i) there is no significant difference between the theoretical gaps of two outer approximations, (ii) the "Gap1" and "Interval Gap" of "ODR-RLB" are smaller than those of "ODR-LB" for most instances. Thus, we recommend that practitioners solve the second outer approximation to obtain a lower bound, where this approximation also provides a theoretical optimality guarantee when $m_1 = K$ (see Theorem 5).

Furthermore, although we obtain near-optimal solutions by setting $m_1 = K$ in the ODR approach, the sensitivity analyses of our ODR approach in Tables A.3–A.5 (see details in Appendix A.7.4) with respect to m_1 also show valuable results. Specifically, we consider the production-transportation problem (2.41), where m_1 takes values from $\{3, 5, 7\}$ when $K = 5$, $\{8, 10, 12\}$ when $K = 10$, and $\{13, 15, 17\}$ when $K = 15$. Concerning our ODR approach (i.e., "ODR-LB," "ODR-UB," and "ODR-RLB"), we observe a general trend where the values of "Gap1," "Gap2," and "Interval Gap" all tend to decrease as m_1 increases. This trend aligns with the theoretical results in Theorems 1, 3, and 5. Meanwhile, we consider the same problem with $G = 5$, $H = 20$, and $K = 10$ and provide a line chart in Figure 2.1 to demonstrate the trend concerning the value of m_1 in a higher granularity. These results demonstrate that (i) the computational times of our three approximations increase when we choose a larger m_1, (ii) for the first outer approximation, the gap between the lower bound and the optimal value decreases from 0.28% to 0.01% when we increase m_1 from 2 to 40, (iii) for the second outer approximation and the inner approximation, a similar conclusion holds, and they will achieve the original optimal value when we choose $m_1 \geq 6$. These results also demonstrate that $m_1 = K$ is a good choice, and increasing m_1 further may not bring significant improvement in terms of solution quality. Combining the tightness results from Theorems 4 and 5 and the sensitivity analyses here, we recommend that practitioners choose $m_1 = K$ to ensure that the proposed approximations yield high-quality results.

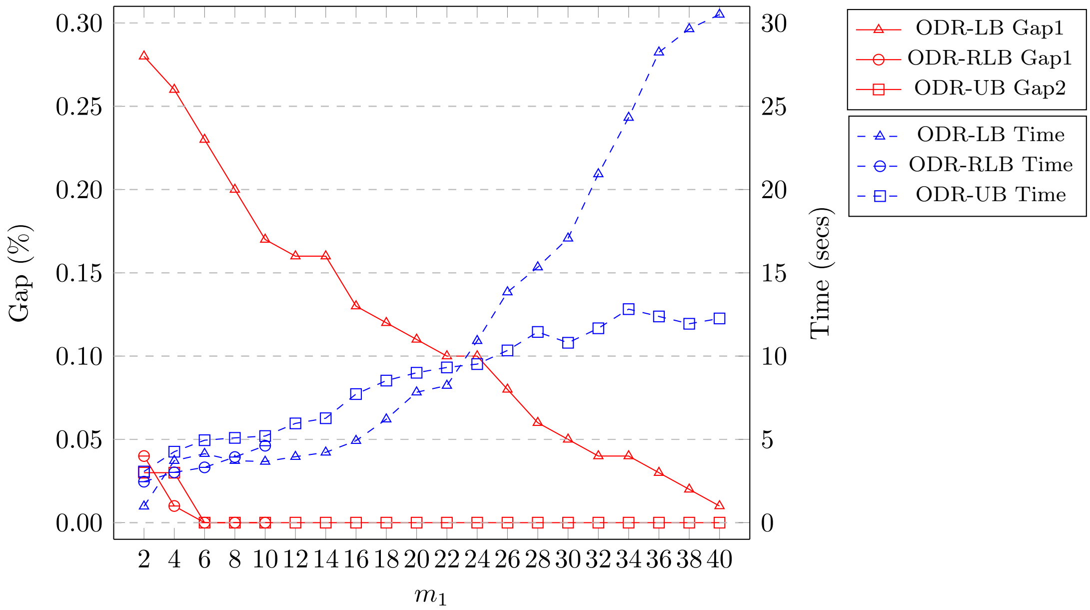

Figure 2.1: Sensitivity Analyses on the Production-Transportation Problem

2.7.2.3 Numerical Insights

In the multiproduct newsvendor problem, Tables 2.1 show that our ODR approach performs better than the PCA approximation with respect to the objective values for all the cases except that the "PCA-100%" (i.e., the original problem) provides the optimal value when the problem size is small. Note that the PCA approximation reduces the dimensionality of the random vector $\boldsymbol{\xi}$ by focusing on only the statistical information of $\boldsymbol{\xi}$, while the ODR approach integrates the dimensionality reduction with the optimization of the original problem. Here we would like to further demonstrate the benefits of our approach, thereby providing insights into how we can choose the value of $\mathbf{B}$ without solving the models in our ODR approach.

Consider the multiproduct newsvendor problem. The PCA approximation chooses the random parameters corresponding to the largest eigenvalues by maximizing the expectation of $\boldsymbol{\xi}^\top\boldsymbol{\xi}$, i.e., the variability of $\boldsymbol{\xi}$. Adopting the idea of our ODR approach to integrate the dimensionality reduction with the subsequent optimization problem, we can consider the objective function $f(\mathbf{x}, \boldsymbol{\xi})$ when choosing the random parameters in $\boldsymbol{\xi}$. Specifically, we can maximize the variability of $(\mathbf{g} - \mathbf{v})^\top\boldsymbol{\xi}$, which is the only random component in

$f(\mathbf{x}, \boldsymbol{\xi})$. By (2.9), we solve the following problem to reduce the dimension from m to m_1:

$$\begin{aligned}
&\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top \boldsymbol{\xi}\boldsymbol{\xi}^\top(\mathbf{g}-\mathbf{v})\right] \\
&\approx \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}+\boldsymbol{\mu}\right)^\top(\mathbf{g}-\mathbf{v})\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top \left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)^\top + 2\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\boldsymbol{\mu}^\top+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)(\mathbf{g}-\mathbf{v})\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[(\mathbf{g}-\mathbf{v})^\top \left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right)^\top +\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)(\mathbf{g}-\mathbf{v})\right].
\end{aligned} \tag{2.43}$$

By introducing $\mathbf{r} = (\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top)(\mathbf{g}-\mathbf{v})$, Problem (2.43) clearly has the same optimal solution as

$$\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \mathbf{r}^\top\mathbf{B}\mathbf{B}^\top\mathbf{r}. \tag{2.44}$$

Proposition 7. *We have* $\mathbf{B}^* = \left[\mathbf{r}/\|\mathbf{r}\|_2, \mathbf{0}_{m\times(m_1-1)}\right]$ *is an optimal solution of Problem* (2.44).

By considering the partial feature of the original optimization problem, the optimal $\mathbf{B}^*$ of Problem (2.43) by Proposition 7 performs better than the PCA approximation that only considers statistical information of random parameters. Note that our proposed inner and outer approximations of the ODR approach consider the complete feature of the original optimization problem and can provide an even better choice of $\mathbf{B}$. In the multi-product newsvendor problem with $K = 2$, we can compare the $\mathbf{B}^*$ of Problem (2.43) with the optimal $\mathbf{B}$ provided by our proposed outer approximation (2.25) with $m_1 = K$. Specifically, letting $m = 10$, we have (i) the optimal value given by the PCA approximation (lower bound) with $m_1 = K$ is -18.62; (ii) the optimal value given by (2.43) is -17.53 with $\mathbf{B} = \begin{bmatrix} -0.8696 & -0.0478 & 0.3285 & -0.0930 & -0.2762 & 0.2126 & -0.0456 & -0.0034 & 0.0361 & 0.0097 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top$; (iii) the optimal value given by (2.25) (lower bound) with $m_1 = K$ is -17.38 with $\mathbf{B} = \begin{bmatrix} -0.8964 & -0.1886 & 0.2094 & -0.0327 & -0.2497 & 0.2215 & -0.0548 & -0.0216 & 0.0289 & 0.0104 \\ 0.0143 & 0.0052 & -0.0014 & -0.0004 & 0.0034 & -0.0035 & 0.0010 & 0.0006 & -0.0003 & -0.0002 \end{bmatrix}^\top$. Clearly, our ODR approach performs the best and the value of $\mathbf{B}$ from solving (2.43) is close to that from our ODR approach (the Frobenius norm of the difference between the two matrices is less than 0.1). That is, if a decision-maker does not have enough capacity to solve the approximations of our ODR approach, the decision-maker may consider the partial feature of the optimization problem when reducing the dimensionality.

2.8 Conclusion

Moment-based DRO provides a theoretical framework to integrate moment-based information from available data with optimal decision-making. Extensive studies have demonstrated the effectiveness of this framework in solving various industrial applications under uncertainties. Although moment-based DRO problems can be reformulated as SDPs that can be solved in polynomial time, solving high-dimensional SDPs is significantly challenging. More importantly, high-dimensional random parameters are generally involved in industrial applications, demanding efficient approaches to solve the high-dimensional SDPs in the context of moment-based DRO.

Current approaches adopt the PCA to first reduce the dimensionality of random parameters using only the statistical information and then solve the subsequent low-dimensional approximation (SDPs). We show that performing dimensionality reduction using the components with the largest variability may not produce a good optimal value from the subsequent PCA approximation and it can be even worse than using the components with the least variability (Example 1). Thus, we integrate the dimensionality reduction with subsequent SDP problems and hence propose an optimized dimensionality reduction (ODR) approach for the moment-based DRO (Sections 2.3–2.5), aiming to drastically reduce the computational time of solving the SDP reformulations while maintaining the optimal solution of the original problem.

We first derive an outer approximation under the ODR approach to provide a lower bound for the optimal value of the original problem (Theorem 1), where the lower bound is nondecreasing in the reduced dimension m_1. We aim to choose a small m_1 to close the approximation gap between the derived lower bound and the original optimal value. To that end, we show that the rank of each SDP matrix with respect to an optimal solution of the original high-dimensional SDP reformulation is small, guiding us on how to optimize the dimensionality reduction (Theorem 2). With this low-rank property, we observe that the derived lower bound can be close to the original optimal value (Theorem 9) but may not reach it (Example 2). Nevertheless, Theorem 9 demonstrates that we are not far from closing the approximation gap and motivates us to derive an inner approximation to provide an upper bound for the optimal value of the original problem (Theorem 3). More importantly, this upper bound reaches the original optimal value when the reduced dimension m_1 is small (Theorem 4). Building on this significant result, we further derive an outer approximation to provide the second lower bound for the optimal value of the original problem, where the gap between the new lower bound and the original optimal value can be closed when the reduced dimension m_1 is small (Theorem 5).

The two outer and one inner approximations derived for the original problem are all low-dimensional SDPs and nonconvex with bilinear terms (Propositions 2 and 3 and Theorem 5). We accordingly develop modified ADMM algorithms to solve them efficiently (Section 2.6) and analyze the convergence property of the ADMM algorithms (see Section 2.6.1). Based on the near-optimal dimensionality reduction solution $\mathbf{B}^{\mathrm{ADMM}}$ returned by the ADMM algorithms, we also explain how to recover the corresponding lower and upper bounds for the original optimal value (see Section 2.6.2). Finally, we demonstrate the effectiveness of our ODR approach in solving multiproduct newsvendor and production-transportation problems. We compare our ODR approach and algorithms with three benchmark approaches: the Mosek solver, the low-rank algorithm by Burer and Monteiro (2003), and existing PCA approximations by Cheramin et al. (2022). The numerical results show that our ODR approach significantly outperforms these benchmarks in computational time and solution quality. Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. More importantly, unlike the existing approaches that become more computationally challenging when the dimension m of random parameters increases, our approach is not sensitive to m, demonstrating significant strength in solving large-scale practical problems (Section 2.7.2.2). In addition, we provide insights into why our ODR approach performs better than the existing PCA approximations (Section 2.7.2.3).

Our research can be further extended in various directions. First, this chapter considers a piece-wise linear cost function in the original problem. Thus, it would be attractive to consider a more general objective function. Second, it would be interesting to apply our approach to more application problems to generate practical insights. Third, it is also of great interest to integrate the idea of dimensionality reduction into the Wasserstein DRO or two-stage stochastic programming. Fourth, our ODR approach can be potentially generalized to solve general SDPs with certain structures. Thus, it would be appealing to exploit the structures of SDP constraints and apply the ODR approach to solve more general SDPs. Fifth, it would be an interesting extension to consider cases with $K \geq m$. Note that for these cases, our proposed ODR approach can still be used to improve the traditional PCA approximation by choosing an appropriate m_1 such that the solution quality is good and the computational time is short, and also provide theoretical guarantees that bound the gap between the proposed upper bound and the original optimal value, as well as the gap between the original optimal value and the proposed lower bound. Nevertheless, we may not provide clear guidance on choosing m_1 so that our proposed approximations achieve the optimal value of the original problem. We leave the above extensions for future research.

Chapter 3

Asymptotically Tight MILP Approximations for a Non-convex QCP

3.1 Introduction

Quadratically constrained programs (QCPs), characterized by quadratic constraints, have attracted substantial attention due to their extensive applications across various industries, including signal processing (Huang and Palomar 2014, De Maio et al. 2010), portfolio optimization (Ban et al. 2018), wireless communications (Zhang and So 2011, Xiang and Tao 2012), and machine learning (Zien and Ong 2007, Bach et al. 2004). For example, the AC optimal power flow (ACOPF) problem in power system operations, as a fundamental industrial problem, has been extensively studied and reformulated as a non-convex QCP (Cain et al. 2012). Similarly, the joint decision and estimation (JDE) problem in signal processing can also be reformulated as a non-convex QCP (Zang and Zhu 2022b). Such significant applications of QCPs motivate us to develop efficient approaches to solve them.

Solving a QCP currently presents a significant challenge. If all quadratic functions are convex, the QCP can be reformulated as a second-order cone program (SOCP), which can be solved in polynomial time. However, including non-convex quadratic functions makes the QCP NP-hard, leading to a computational challenge. Various optimization solvers (Sahinidis 1996, Tawarmalani and Sahinidis 2004, Belotti et al. 2009, Lin and Schrage 2009, Misener and Floudas 2013, 2014, Bestuzheva et al. 2021) have been developed to tackle non-convex QCPs. Among them, the Gurobi solver (Gurobi 2024) currently stands as one of the most widely used tools for solving non-convex QCPs. Specifically, a spatial branch-and-bound framework (Al-Khayyal et al. 1995, Audet et al. 2000, Linderoth 2005) is employed to search for the optimal solution. At each node of the branch-and-bound tree, a relaxation of the subproblem needs to be solved, providing a dual bound for bounding and pruning. The quality of the relaxation significantly impacts computational time, as a tighter relaxation can substantially reduce the search time in the branch-and-bound tree.

Current relaxation techniques primarily involve McCormick inequalities (Misener and Floudas 2012), reformulation-linearization technique (RLT) (Sherali and Adams 2013), semidefinite programming (SDP) (Burer and Vandenbussche 2008), and various combinations thereof. McCormick inequalities typically yield a linear programming relaxation at each node, which is easy to solve but provides a loose dual bound. SDP relaxation may offer a tighter dual bound but is time-consuming to solve. RLT can be employed to further tighten existing relaxations. Shor relaxation (Shor 1987) represents a fundamental SDP relaxation for general QCPs, introducing the variable $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and relaxing it to $\mathbf{X} \succeq \mathbf{x}\mathbf{x}^\top$. Building upon these foundations, Sturm and Zhang (2003) and Burer and Yang (2015) combine SDP relaxations with the RLT (Sherali and Adams 2013) to achieve exact relaxations for specific classes of QCPs. A comprehensive survey by Bao et al. (2011) conducts an in-depth comparison of existing SDP relaxations for QCPs. This survey covers various approaches, including standard Lagrangian relaxation (Wolkowicz 2000) and its variants (Poljak et al. 1995, Shor 1992, Wolkowicz 2000), Shor relaxation (Shor 1990) and its variants (Sherali 2007), and doubly nonnegative relaxation (Burer

2009). Specifically, Bao et al. (2011) demonstrate that the doubly nonnegative relaxation and a specific variant of the Shor relaxation may outperform others in terms of relaxation gap. In addition, other convex relaxations for non-convex QCPs are also studied. For example, Jiang and Li (2019) use RLT to tighten the second-order cone (SOC)-constrained convex relaxations of non-convex QCPs.

In this chapter, we present a novel approach to developing tighter relaxations for a non-convex QCP using *mixed-integer linear programming (MILP) approximations*. The goal is to provide a small optimality gap for the non-convex QCP within a short computational time. To the best of our knowledge, we are among the first to utilize MILP approximations to solve non-convex QCPs, thereby providing an advantageous method to leverage state-of-the-art MILP solvers. One relevant research is by Xia et al. (2020), which provides an MILP reformulation for non-convex quadratic programs with linear constraints. Another relevant research in this context is by Dong and Luo (2018), which primarily focuses on general mixed-integer quadratically constrained programs (MIQCPs). The authors convexify all the non-convex quadratic constraints using diagonal perturbation and consolidate the non-convexity within the constraint $y = x^2$. They then approximate the relationship $y = x^2$, eventually relaxing the MIQCP into a mixed-integer SOCP (MISOCP). In our work, we develop a different approach by decomposing a non-convex quadratic constraint into two constraints: $\|\mathbf{y}\| \leq t$ and $\|\mathbf{z}\| \geq t$, where $\mathbf{y}$, $\mathbf{z}$, and t will be properly defined. Such a decomposition allows us to incorporate the non-convexity of the entire problem within the constraint $\|\mathbf{z}\| \geq t$. Subsequently, we employ two methods to approximate $\|\mathbf{y}\| \leq t$ and $\|\mathbf{z}\| \geq t$, thereby approximating the non-convex QCP with two MILPs. The key contributions of our research can be summarized as follows:

- We first utilize an eigenvalue-based decomposition to express a non-convex quadratic function as the difference of two convex functions. We then introduce an additional variable t to partition each non-convex quadratic constraint into an SOC constraint $\|\mathbf{y}\| \leq t$ and the complement of an SOC constraint $\|\mathbf{z}\| \geq t$. To approximate

$\|\mathbf{y}\| \leq t$, we employ two polyhedral approximation approaches. The constraint $\|\mathbf{z}\| \geq t$ is approximated using a combination of linear and complementarity constraints. That is, we approximate the non-convex QCP with two linear programs with complementarity constraints (LPCCs). Generally, the first LPCC approximation involves more variables, while the second has more constraints. We investigate the trade-off between them through computational experiments.

- We demonstrate that the optimal values of the LPCCs asymptotically converge to the optimal value of the original non-convex QCP as the approximation precision increases. In addition, we establish the boundedness of the LPCCs by leveraging the boundedness property of the original non-convex QCP. This boundedness allows the LPCCs to be reformulated as MILPs. To the best of our knowledge, we are among the first to propose LPCC/MILP approximations for non-convex QCPs.

- We perform extensive experiments to validate the effectiveness of our proposed approximations, which yield smaller optimality gaps within shorter computational times compared to a state-of-the-art commercial solver. To demonstrate the practicality of our approach, we consider two applications: the joint decision and estimation (JDE) problem and the two-trust-region subproblem (TTRS). The numerical results substantiate the advantages of our proposed approximations in terms of optimality gap and computational time.

The remainder of this chapter is organized as follows. In Section 3.2, we approximate a non-convex QCP with an LPCC and present theoretical results for this LPCC, which can be further reformulated as an MILP. In Section 3.3, we propose an alternative MILP approximation using a different polyhedral approximation of the SOC constraint. In Section 3.4, we perform extensive experiments on our MILP approximations and provide two applications. In Section 3.5, we conclude this chapter.

Notation

For any integer $n \geq 1$, we use $[n]$ to represent the set $\{1, 2, \ldots, n\}$. Vectors in n-dimensional Euclidean space $\mathbb{R}^n$ are denoted by bold-faced lowercase letters (e.g., zero vector $\mathbf{0}$), and matrices are denoted by bold-faced uppercase letters (e.g., zero matrix $\mathbf{O}$). The symbol $^\top$ denotes transposition. For any vector $\mathbf{x} \in \mathbb{R}^n$, we denote $|\mathbf{x}| = [|x_1|, \ldots, |x_n|]^\top$. For any matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, $|\mathbf{X}|$ represents the matrix with elementwise absolute values of $\mathbf{X}$. We let the symbol $\|\cdot\|$ denote the L2-norm. The nonnegative orthant is denoted by $\mathbb{R}^n_+ := \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0, \ \forall\, i \in [n]\}$. The notation $\mathbf{I}_n$ represents the $n \times n$ identity matrix with columns $\mathbf{e}_i$ for all $i \in [n]$, while $\mathbf{e} := \sum_{i=1}^{n} \mathbf{e}_i = [1, \ldots, 1]^\top \in \mathbb{R}^n$. We use $\mathbb{R}^{d \times p}$ and $\mathbb{R}^{d \times p}_+$ to denote the set of all $d \times p$ matrices and the subset of those with nonnegative entries, respectively. For a given symmetric matrix $\mathbf{H} = \mathbf{H}^\top$, $\mathbf{H} \succeq 0$ indicates that $\mathbf{H}$ is positive semidefinite (PSD). We let $\mathcal{S}^d := \{\mathbf{X} \in \mathbb{R}^{d \times d} \mid \mathbf{X} = \mathbf{X}^\top\}$ be the set of $d \times d$ symmetric matrices, and $\mathbb{Z}_{++}$ denote the set of all strictly positive integers.

3.2 Approximation of QCP

In this section, we present our approach to approximating a non-convex QCP. We decompose each non-convex quadratic constraint in the non-convex QCP into an SOC constraint and the complement of an SOC constraint. We introduce a method for approximating these components and then approximate the non-convex QCP with an LPCC. We demonstrate that the optimal value of the LPCC asymptotically converges to that of the QCP. We further establish the boundedness of the feasible region for the LPCC, which allows us to reformulate it into an MILP.

3.2.1 Decomposition of Quadratic Constraints

We consider the following non-convex QCP:

$$\Theta^* := \min_{\mathbf{x}} \ \ \mathbf{c}_0^\top \mathbf{x} \tag{3.1}$$

$$\text{s.t.}\ \ \mathbf{x}^\top \mathbf{Q}_i \mathbf{x} + 2\mathbf{c}_i^\top \mathbf{x} \le b_i,\ \forall\, i \in [m],$$

where $\mathbf{Q}_i \in \mathcal{S}^n$ for all $i \in [m]$, $\mathbf{c}_i \in \mathbb{R}^n$ for all $i \in \{0, 1, \dots, m\}$, and $\mathbf{b} := [b_1, \dots, b_m]^\top \in \mathbb{R}^m$. Throughout this chapter, we assume that the feasible region of Problem (3.1) is bounded. That is, $\|\mathbf{x}\| \le U$, where $U > 0$. Note that Problem (3.1) includes more general quadratically constrained quadratic programs because we can make the objective function linear by considering the epigraph form of the quadratically constrained quadratic program.

For each $i \in [m]$, we let $\mathbf{Q}_i = \mathbf{A}_i \mathbf{A}_i^\top - \mathbf{B}_i \mathbf{B}_i^\top$, where $\mathbf{A}_i \in \mathbb{R}^{n \times k_i^+}$, $\mathbf{B}_i \in \mathbb{R}^{n \times k_i^-}$, k_i^+ is the number of nonnegative eigenvalues of $\mathbf{Q}_i$, and k_i^- is the number of negative eigenvalues of $\mathbf{Q}_i$. Thus, Problem (3.1) becomes

$$\begin{aligned} \Theta^* = \min_{\mathbf{x}} \quad & \mathbf{c}_0^\top \mathbf{x} && (3.2a) \\ \text{s.t.} \quad & (\mathbf{A}_i^\top \mathbf{x})^\top \mathbf{A}_i^\top \mathbf{x} - (\mathbf{B}_i^\top \mathbf{x})^\top \mathbf{B}_i^\top \mathbf{x} + 2\mathbf{c}_i^\top \mathbf{x} \le b_i,\ \forall\, i \in [m]. && (3.2b) \end{aligned}$$

By the definition of L2-norm, for any $i \in [m]$, we have

$$\begin{aligned} (3.2b) \iff & \left\|\mathbf{A}_i^\top \mathbf{x}\right\|^2 - \left\|\mathbf{B}_i^\top \mathbf{x}\right\|^2 \le \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right)^2 - \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right)^2 \\ \iff & \left\|\mathbf{A}_i^\top \mathbf{x}\right\|^2 + \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right)^2 \le \left\|\mathbf{B}_i^\top \mathbf{x}\right\|^2 + \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right)^2 \\ \iff & \left\|\left[\mathbf{A}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right]\right\|^2 \le \left\|\left[\mathbf{B}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right]\right\|^2 \\ \iff & \left\|\left[\mathbf{A}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right]\right\| \le \left\|\left[\mathbf{B}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right]\right\|. \end{aligned}$$

By inserting auxiliary variables between two norms, Problem (3.2) can be further reformulated as

$$\begin{aligned} \Theta^* = \min_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{t}} \quad & \mathbf{c}_0^\top \mathbf{x} && (3.3a) \\ \text{s.t.} \quad & \|\mathbf{y}_i\| \le t_i,\ \forall\, i \in [m], && (3.3b) \end{aligned}$$

$$\|\mathbf{z}_i\| \geq t_i, \ \forall\, i \in [m], \tag{3.3c}$$

$$\mathbf{y}_i = \left[\mathbf{A}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right] \in \mathbb{R}^{k_i^+ + 1}, \ \forall\, i \in [m], \tag{3.3d}$$

$$\mathbf{z}_i = \left[\mathbf{B}_i^\top \mathbf{x}; \frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right] \in \mathbb{R}^{k_i^- + 1}, \ \forall\, i \in [m], \tag{3.3e}$$

where $\mathbf{y}_i, \mathbf{z}_i$, and t_i are auxiliary variables for all $i \in [m]$. Under the boundedness assumption of $\mathbf{x}$, we have that $\mathbf{y}_i$ and $\mathbf{z}_i$ are also bounded for all $i \in [m]$. That is, there exists $U_y > 0$ and $U_z > 0$ such that $\|\mathbf{y}_i\| \leq U_y$ and $\|\mathbf{z}_i\| \leq U_z$ for all $i \in [m]$ (see Appendix B.2 for further details of U_y and U_z). We approximate $\|\mathbf{y}_i\| \leq t_i$ and $\|\mathbf{z}_i\| \geq t_i$ for each $i \in [m]$ in the following Sections 3.2.2 and 3.2.3, respectively.

3.2.2 Polyhedral Approximation of $\|\mathbf{y}\| \leq t$

Given $\|\mathbf{y}_i\| \leq t_i$ for all $i \in [m]$ in (3.3b), here we omit the subscript i and consider the SOC constraint $\|\mathbf{y}\| \leq t$ with $\mathbf{y} \in \mathbb{R}^k$, $k \in \mathbb{Z}_{++}$, and $t \in \mathbb{R}$. We let

$$L^k := \left\{(\mathbf{y}, t) \in \mathbb{R}^k \times \mathbb{R} \;\middle|\; \|\mathbf{y}\| \leq t\right\}$$

be the $(k+1)$-dimensional SOC. Given an approximation quality parameter $\varepsilon \in (0, 1]$, Ben-Tal and Nemirovski (2001) construct a polyhedral ε-approximation of L^k

$$\mathcal{P}_\varepsilon^k := \left\{(\mathbf{y}, t) \in \mathbb{R}^k \times \mathbb{R} \;\middle|\; \mathbf{P}_\varepsilon^k \mathbf{y} + \mathbf{H}_\varepsilon^k \mathbf{u} + \mathbf{p}_\varepsilon^k t \geq \mathbf{0} \text{ for some } \mathbf{u}\right\}$$

such that $L^k \subseteq \mathcal{P}_\varepsilon^k$ and if $(\mathbf{y}, t) \in \mathcal{P}_\varepsilon^k$, then $\|\mathbf{y}\| \leq (1+\varepsilon)t$.

To derive the detailed formulation of $\mathcal{P}_\varepsilon^k$, we first represent $\|\mathbf{y}\| \leq t$ using a system of tower constraints. We let θ be the smallest integer such that $k \leq 2^\theta$ and use y_i^ℓ to denote the ith variable of level ℓ of the tower for any $\ell \in [\theta]$ and $i \in [2^{\theta-\ell}]$. Specifically, $y_1^0, \ldots, y_k^0$ represent the original variables in the vector $\mathbf{y}$, $y_1^\theta = t$, and variables $y_{2i-1}^{\ell-1}$ and $y_{2i}^{\ell-1}$ are the "parents" of y_i^ℓ for any $\ell \in [\theta]$. Note that $y_{k+1}^0, \ldots, y_{2^\theta}^0$ and y_i^l ($\forall i \in [2^{\theta-\ell}], \ell \in [\theta-1]$) are extra variables and not included in the vector $\mathbf{y}$. Using this system of tower constraints, we represent L^k as

$$L^k = \left\{ (\mathbf{y}, t) \in \mathbb{R}^k \times \mathbb{R} \;\middle|\; \begin{array}{l} \sqrt{\left(y_{2i-1}^{\ell-1}\right)^2 + \left(y_{2i}^{\ell-1}\right)^2} \leq y_i^\ell, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta] \\ \text{for some } y_{k+1}^0, \ldots, y_{2^\theta}^0, \text{ and } y_i^\ell, \; \forall i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta-1] \end{array} \right\}. \tag{3.4}$$

Then, for each $\ell \in [\theta]$ and $i \in [2^{\theta-\ell}]$, we use linear constraints to approximate $\sqrt{(y_{2i-1}^{\ell-1})^2 + (y_{2i}^{\ell-1})^2} \leq y_i^\ell$. In the end, $\mathcal{P}_\varepsilon^k$ is defined by the following linear constraints:

$$\xi_{\ell,i}^0 \geq y_{2i-1}^{\ell-1}, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5a}$$

$$\xi_{\ell,i}^0 \geq -y_{2i-1}^{\ell-1}, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5b}$$

$$\eta_{\ell,i}^0 \geq y_{2i}^{\ell-1}, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5c}$$

$$\eta_{\ell,i}^0 \geq -y_{2i}^{\ell-1}, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5d}$$

$$\xi_{\ell,i}^j = \cos\left(\frac{\pi}{2^{j+1}}\right) \xi_{\ell,i}^{j-1} + \sin\left(\frac{\pi}{2^{j+1}}\right) \eta_{\ell,i}^{j-1}, \; \forall\, j \in [\nu_\ell], \; i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5e}$$

$$\eta_{\ell,i}^j \geq -\sin\left(\frac{\pi}{2^{j+1}}\right) \xi_{\ell,i}^{j-1} + \cos\left(\frac{\pi}{2^{j+1}}\right) \eta_{\ell,i}^{j-1}, \; \forall\, j \in [\nu_\ell], \; i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5f}$$

$$\eta_{\ell,i}^j \geq \sin\left(\frac{\pi}{2^{j+1}}\right) \xi_{\ell,i}^{j-1} - \cos\left(\frac{\pi}{2^{j+1}}\right) \eta_{\ell,i}^{j-1}, \; \forall\, j \in [\nu_\ell], \; i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5g}$$

$$\xi_{\ell,i}^{\nu_\ell} \leq y_i^\ell, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5h}$$

$$\eta_{\ell,i}^{\nu_\ell} \leq \tan\left(\frac{\pi}{2^{\nu_\ell+1}}\right) \xi_{\ell,i}^{\nu_\ell}, \; \forall\, i \in \left[2^{\theta-\ell}\right], \; \ell \in [\theta], \tag{3.5i}$$

where $y_i^0 = y_i$ for all $i \in [k]$, $y_1^\theta = t$, and the vector $\mathbf{u}$ in the simplified definition of $\mathcal{P}_\varepsilon^k$ is the collection of all extra variables $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, $y_{k+1}^0, \ldots, y_{2^\theta}^0$, and y_i^ℓ ($\forall i \in [2^{\theta-\ell}], \ell \in [\theta-1]$). Note that $\nu_\ell \in \mathbb{Z}_{++}$ ($\forall \ell \in [\theta]$) are parameters used to control the quality of the polyhedral approximation. Specifically, as demonstrated by Ben-Tal and Nemirovski (2001), the overall quality ε can be expressed as

$$\varepsilon = \prod_{\ell=1}^{\theta} (1 + \varepsilon_\ell) - 1, \tag{3.6}$$

where ε_ℓ is the approximation quality at level ℓ, defined as

$$\varepsilon_\ell = \frac{1}{\cos\left(\frac{\pi}{2^{\nu_\ell+1}}\right)} - 1. \tag{3.7}$$

It is worth noting that the number of additional variables and linear constraints in this approximation is in the linear order of $k\ln(2/\varepsilon)$.

Remark 1. *Given any $\varepsilon \in (0,1]$ and $\theta \in \mathbb{Z}_{++}$, $\mathcal{P}_\varepsilon^{k_1}$ and $\mathcal{P}_\varepsilon^{k_2}$ share the same formulation for any k_1 and k_2 that belong to $\{2^{\theta-1}+1,\dots,2^\theta\}$. Therefore, without loss of generality, we assume $k = 2^\theta$ for some $\theta \in \mathbb{Z}_{++}$ such that $\mathbf{y} \in \mathbb{R}^k$ in the set L^k.*

Remark 2. *Given that $\mathbf{y}$ and t are bounded due to the setting in Section 3.2.1, it is clear that $\mathcal{P}_\varepsilon^k$ is bounded, implying that the variable $\mathbf{u}$ is bounded.*

3.2.3 Approximation of $\|\mathbf{z}\| \geq t$

Given $\|\mathbf{z}_i\| \geq t_i$ for all $i \in [m]$ in (3.3c), we omit the subscript i and consider $\|\mathbf{z}\| \geq t$ with $\mathbf{z} \in \mathbb{R}^k$, $k \in \mathbb{Z}_{++}$, and $t \in \mathbb{R}$. We let $\bar{L}^k := \{(\mathbf{z}, t) \in \mathbb{R}^k \times \mathbb{R} \mid \|\mathbf{z}\| \geq t\}$ be the complement of the $(k+1)$-dimensional SOC L^k. By the definition of Euclidean norm, $\|\mathbf{z}\| = \max_{\mathbf{w}}\{\mathbf{w}^\top\mathbf{z} \mid \|\mathbf{w}\| \leq 1\}$. As discussed in Section 3.2.2, we have

$$\begin{aligned}\|\mathbf{z}\| \geq t \iff & \exists\, \mathbf{w} \text{ s.t. } \|\mathbf{w}\| \leq 1 \text{ and } \mathbf{w}^\top\mathbf{z} \geq t \\ \implies & \exists\, \mathbf{w} \text{ s.t. } (\mathbf{w}, 1) \in \mathcal{P}_\varepsilon^k \text{ and } \mathbf{w}^\top\mathbf{z} \geq t.\end{aligned}$$

As a result, $\bar{L}^k$ can be approximated by the set $G := \{(\mathbf{z}, t) \mid (\mathbf{w}, 1) \in \mathcal{P}_\varepsilon^k,\ \mathbf{w}^\top\mathbf{z} \geq t \text{ for some } \mathbf{w}\}$.

We then represent G by linear and complementarity constraints. Specifically, we have the following proposition:

Proposition 8. $(\mathbf{z}, t) \in G$ *if and only if*

$$
\begin{aligned}
t \leq \max_{\mathbf{w},\mathbf{s}} \quad & \mathbf{w}^\top \mathbf{z} && \text{(3.8a)}\\
s.t. \quad & \mathbf{P}^k_\varepsilon \mathbf{w} + \mathbf{H}^k_\varepsilon \mathbf{s} + \mathbf{p}^k_\varepsilon \geq \mathbf{0}. && \text{(3.8b)}
\end{aligned}
$$

Proof. By definition of G, we have

$$
\begin{aligned}
G :=& \left\{(\mathbf{z}, t) \mid (\mathbf{w}, 1) \in \mathcal{P}^k_\varepsilon,\ \mathbf{w}^\top \mathbf{z} \geq t \text{ for some } \mathbf{w}\right\}\\
=& \left\{(\mathbf{z}, t) \mid \mathbf{P}^k_\varepsilon \mathbf{w} + \mathbf{H}^k_\varepsilon \mathbf{s} + \mathbf{p}^k_\varepsilon \geq \mathbf{0},\ \mathbf{w}^\top \mathbf{z} \geq t \text{ for some } \mathbf{w}, \mathbf{s}\right\}\\
=& \left\{(\mathbf{z}, t) \mid \max_{\mathbf{w},\mathbf{s}} \left\{\mathbf{w}^\top \mathbf{z} \mid \mathbf{P}^k_\varepsilon \mathbf{w} + \mathbf{H}^k_\varepsilon \mathbf{s} + \mathbf{p}^k_\varepsilon \geq \mathbf{0}\right\} \geq t\right\},
\end{aligned}
$$

where the second equality holds by the definition of $\mathcal{P}^k_\varepsilon$ and the last equality holds by the equivalence between "for some" and "max". □

Equivalently, $(\mathbf{z}, t) \in G$ if and only if

$$
\begin{aligned}
t \leq \min_{\mathbf{v}} \quad & (\mathbf{p}^k_\varepsilon)^\top \mathbf{v} && \text{(3.9a)}\\
\text{s.t.} \quad & (\mathbf{P}^k_\varepsilon)^\top \mathbf{v} = -\mathbf{z}, && \text{(3.9b)}\\
& (\mathbf{H}^k_\varepsilon)^\top \mathbf{v} = \mathbf{0}, && \text{(3.9c)}\\
& \mathbf{v} \geq \mathbf{0}, && \text{(3.9d)}\\
& (\mathbf{p}^k_\varepsilon)^\top \mathbf{v} \leq M, && \text{(3.9e)}
\end{aligned}
$$

where $M > 0$ is a large enough number. Here (3.9e) is a redundant constraint to bound the objective function $(\mathbf{p}^k_\varepsilon)^\top \mathbf{v}$. Note that (3.8) is equivalent to (3.9a)–(3.9d) by the strong duality of linear programming. Since the vectors $\mathbf{w}$ and $\mathbf{z}$ in (3.8) are bounded, the optimal value of the optimization problem $\min_{\mathbf{v}}\{(\mathbf{p}^k_\varepsilon)^\top \mathbf{v} \mid \text{(3.9b)} - \text{(3.9d)}\}$ is bounded. Therefore, we can claim that (3.9e) is a redundant constraint. This redundant constraint will be useful in Section 3.2.5.

Remark 3. *In* (3.9e)*, we can set* $M = (1+\varepsilon)U_z$ *because* $\|\mathbf{w}\| \leq 1+\varepsilon$*,* $\|\mathbf{z}\| \leq U_z$*, and* $\mathbf{w}^\top \mathbf{z} \leq \|\mathbf{w}\|\|\mathbf{z}\| \leq (1+\varepsilon)U_z$*.*

By the KKT conditions of linear programming, condition (3.9) can be further represented by the following linear and complementarity constraints:

$$(\mathbf{p}_\varepsilon^k)^\top \mathbf{v} - t \geq 0, \tag{3.10a}$$

$$\mathbf{P}_\varepsilon^k \mathbf{w} + \mathbf{H}_\varepsilon^k \mathbf{s} \geq -\mathbf{p}_\varepsilon^k, \tag{3.10b}$$

$$(\mathbf{P}_\varepsilon^k)^\top \mathbf{v} + \mathbf{z} = \mathbf{0}, \tag{3.10c}$$

$$(\mathbf{H}_\varepsilon^k)^\top \mathbf{v} = \mathbf{0}, \tag{3.10d}$$

$$\mathbf{v} \geq \mathbf{0}, \tag{3.10e}$$

$$(\mathbf{p}_\varepsilon^k)^\top \mathbf{v} \leq M, \tag{3.10f}$$

$$\mathbf{v}^\top \left(\mathbf{P}_\varepsilon^k \mathbf{w} + \mathbf{H}_\varepsilon^k \mathbf{s} + \mathbf{p}_\varepsilon^k\right) = 0. \tag{3.10g}$$

To summarize, we have

$$\bar{L}^k \subseteq \left\{(\mathbf{z}, t) \in \mathbb{R}^k \times \mathbb{R} \mid (3.10) \text{ for some } \mathbf{w}, \mathbf{s}, \mathbf{v}\right\}.$$

Remark 4. *In* (3.10)*, it is clear that variables* $\mathbf{z}$*,* t*,* $\mathbf{w}$*, and* $\mathbf{s}$ *are bounded (see Appendix B.2 and Remark 2). However, the variable* $\mathbf{v}$ *may be unbounded because it is the dual variable of an LP. Thus, we add the redundant constraint* (3.10f)*, which will be useful later in proving the boundedness of* $\mathbf{v}$*.*

3.2.4 LPCC Approximation

With the approximation results in Sections 3.2.2 and 3.2.3, for any $\varepsilon \in (0,1]$, we can now approximate Problem (3.3) with the following LPCC:

$$
\begin{aligned}
\Theta_\varepsilon := \min_{\substack{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\\ \mathbf{v},\mathbf{w},\mathbf{s},\mathbf{t}}} \quad & \mathbf{c}_0^\top \mathbf{x} && \text{(3.11a)}\\
\text{s.t.} \quad & \mathbf{P}_\varepsilon^{k_i^+ +1}\mathbf{y}_i + \mathbf{H}_\varepsilon^{k_i^+ +1}\mathbf{u}_i + \mathbf{p}_\varepsilon^{k_i^+ +1} t_i \geq \mathbf{0},\ \forall i \in [m], && \text{(3.11b)}\\
& \left(\mathbf{p}_\varepsilon^{k_i^- +1}\right)^\top \mathbf{v}_i - t_i \geq 0,\ \forall i \in [m], && \text{(3.11c)}\\
& \mathbf{P}_\varepsilon^{k_i^- +1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^- +1}\mathbf{s}_i \geq -\mathbf{p}_\varepsilon^{k_i^- +1},\ \forall i \in [m], && \text{(3.11d)}\\
& \left(\mathbf{P}_\varepsilon^{k_i^- +1}\right)^\top \mathbf{v}_i + \mathbf{z}_i = \mathbf{0},\ \forall i \in [m], && \text{(3.11e)}\\
& \left(\mathbf{H}_\varepsilon^{k_i^- +1}\right)^\top \mathbf{v}_i = \mathbf{0},\ \forall i \in [m], && \text{(3.11f)}\\
& \mathbf{v}_i \geq \mathbf{0},\ \forall i \in [m], && \text{(3.11g)}\\
& \left(\mathbf{p}_\varepsilon^{k_i^- +1}\right)^\top \mathbf{v}_i \leq M,\ \forall i \in [m], && \text{(3.11h)}\\
& \mathbf{v}_i^\top \left(\mathbf{P}_\varepsilon^{k_i^- +1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^- +1}\mathbf{s}_i + \mathbf{p}_\varepsilon^{k_i^- +1}\right) = 0,\ \forall i \in [m], && \text{(3.11i)}\\
& \text{(3.3d)} - \text{(3.3e)}. && \text{(3.11j)}
\end{aligned}
$$

Clearly, the optimal value of the LPCC described above provides a lower bound for the optimal value of Problem (3.1).

Subsequently, we prove that the optimal value of the LPCC asymptotically converges to that of the original non-convex QCP, i.e., Problem (3.1). This result is based on the following proposition.

Proposition 9. *If* $(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{w},\mathbf{s},\mathbf{t})$ *is feasible for* (3.11)*, then* $(\mathbf{x},\mathbf{y},\mathbf{z})$ *is feasible to the following "ε-relaxation" of Problem* (3.1)*:*

$$
\begin{aligned}
\underline{\Theta}_\varepsilon := \min_{\mathbf{x},\mathbf{y},\mathbf{z}} \quad & \mathbf{c}_0^\top \mathbf{x} && \text{(3.12a)}\\
s.t. \quad & \|\mathbf{y}_i\| \leq (1+\varepsilon)^2 \|\mathbf{z}_i\|,\ \forall i \in [m], && \text{(3.12b)}\\
& \text{(3.3d)} - \text{(3.3e)}. && \text{(3.12c)}
\end{aligned}
$$

Proof. We prove that the feasible set of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ induced by constraints (3.11b)–(3.11i) is a subset of the feasible region induced by constraints (3.12b). Note that Problem (3.11) is derived using the approximations to Problem (3.3) in Sections 3.2.2 and 3.2.3. Thus, for any $i \in [m]$, if constraints (3.11b) are satisfied, that is, $(\mathbf{y}_i, t_i) \in \mathcal{P}_\varepsilon^{k_i^+ + 1}$, then $\|\mathbf{y}_i\| \leq (1+\varepsilon)t_i$. If constraints (3.11c)–(3.11i) are satisfied, then

$$t_i \leq \max_{\mathbf{w}} \left\{ \mathbf{w}^\top \mathbf{z}_i \;\middle|\; (\mathbf{w}, 1) \in \mathcal{P}_\varepsilon^{k_i^- + 1} \right\} \leq \max_{\mathbf{w}} \left\{ \mathbf{w}^\top \mathbf{z}_i \;\middle|\; \|\mathbf{w}\| \leq 1 + \varepsilon \right\} = (1+\varepsilon)\|\mathbf{z}_i\|.$$

Therefore, if constraints (3.11b)–(3.11i) are satisfied, then $\|\mathbf{y}_i\| \leq (1+\varepsilon)^2 \|\mathbf{z}_i\|$. $\square$

Proposition 9 indicates that the violation of constraints approaches zero as ε tends to zero. Based on this proposition, we derive the following asymptotic convergence conclusion for the LPCC.

Theorem 7. *We have* $\lim_{\varepsilon \to 0^+} \Theta_\varepsilon = \Theta^*$.

Proof. By Proposition 9, we have $\underline{\Theta}_\varepsilon \leq \Theta_\varepsilon \leq \Theta^*$ for any $\varepsilon \in (0, 1]$. Therefore, it suffices to show that $\lim_{\varepsilon \to 0^+} \underline{\Theta}_\varepsilon = \Theta^*$.

It is clear that $\underline{\Theta}_{\varepsilon_1} \leq \underline{\Theta}_{\varepsilon_2} \leq \Theta^*$ for any $0 < \varepsilon_2 \leq \varepsilon_1 \leq 1$. Thus, $\lim_{\varepsilon \to 0^+} \underline{\Theta}_\varepsilon$ exists and $\lim_{\varepsilon \to 0^+} \underline{\Theta}_\varepsilon \leq \Theta^*$. For any $\varepsilon = 1/k$ with $k \in \mathbb{Z}_{++}$, let $(\mathbf{x}_k^*, \mathbf{y}_k^*, \mathbf{z}_k^*)$ be an optimal solution of (3.12). Since $\{\mathbf{x}_k^*\}$ is a bounded sequence, it has a convergent subsequence $\{\mathbf{x}_{k_j}^*\}$ such that $k_j \to \infty$ as $j \to \infty$ and $\lim_{j \to \infty} \mathbf{x}_{k_j}^*$ exists. Denote the limit by $\bar{\mathbf{x}}$. Then, we have

$$\mathbf{c}^\top \bar{\mathbf{x}} = \lim_{j \to \infty} \mathbf{c}^\top \mathbf{x}_{k_j}^* = \lim_{j \to \infty} \underline{\Theta}_{\frac{1}{k_j}} = \lim_{\varepsilon \to 0^+} \underline{\Theta}_\varepsilon.$$

In addition, since $(\mathbf{x}_{k_j}^*, \mathbf{y}_{k_j}^*, \mathbf{z}_{k_j}^*)$ is feasible to (3.12) for each j, it follows that

$$\mathbf{x}_{k_j}^\top \mathbf{Q}_i \mathbf{x}_{k_j} + 2\mathbf{c}_i^\top \mathbf{x}_{k_j} - b_i \leq \frac{1}{k_j}\left(\frac{1}{k_j} + 2\right)\left(\mathbf{x}_{k_j}^\top \mathbf{B}_i \mathbf{B}_i^\top \mathbf{x}_{k_j} + \frac{\left(b_i - 2\mathbf{c}_i^\top \mathbf{x}_{k_j} + 1\right)^2}{4}\right), \ \forall i \in [m].$$

As $j \to \infty$, we have $\bar{\mathbf{x}}^\top \mathbf{Q}_i \bar{\mathbf{x}} + 2\mathbf{c}_i^\top \bar{\mathbf{x}} - b_i \leq 0$ for all $i \in [m]$. That is, $\bar{\mathbf{x}}$ is feasible to

Problem (3.1). Therefore, $\lim_{\varepsilon\to 0^+} \underline{\Theta}_\varepsilon = \mathbf{c}^\top \bar{\mathbf{x}} \geq \Theta^*$, which completes the proof. $\square$

3.2.5 MILP Reformulation

In Section 3.2.4, we approximate a non-convex QCP (i.e., Problem (3.1)) with an LPCC (i.e., Problem (3.11)). This approach enables us to utilize all existing and further developed LPCC solvers to solve the approximation of QCP. Currently, MILP solvers are more mature and efficient than LPCC solvers. Therefore, to leverage the efficiency of current MILP solvers, we transform the LPCC into an MILP. Specifically, we employ the Big-M method to reformulate (3.11i) into mixed-integer linear constraints. To achieve this, we need to establish the boundedness of $\mathbf{v}_i$ ($\forall i \in [m]$) and $\mathbf{P}_\varepsilon^{k_i^-+1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^-+1}\mathbf{s}_i + \mathbf{p}_\varepsilon^{k_i^-+1}$ ($\forall i \in [m]$).

By the construction of the polyhedral approximation in Section 3.2.2 and the approximation in Section 3.2.3, for any $i \in [m]$, we have

$$\|\mathbf{w}_i\| \leq 1+\varepsilon,\ |\mathbf{s}_i| \leq (1+\varepsilon)\mathbf{e}.$$

Given that $\mathbf{w}_i$ and $\mathbf{s}_i$ are bounded for each $i \in [m]$, it is clear that $\mathbf{P}_\varepsilon^{k_i^-+1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^-+1}\mathbf{s}_i + \mathbf{p}_\varepsilon^{k_i^-+1}$ is bounded from above by $(1+\varepsilon)|\mathbf{P}_\varepsilon^{k_i^-+1}|\mathbf{e} + (1+\varepsilon)|\mathbf{H}_\varepsilon^{k_i^-+1}|\mathbf{e} + |\mathbf{p}_\varepsilon^{k_i^-+1}|$. It follows that we only need to confirm the boundedness of $\mathbf{v}_i$ ($\forall i \in [m]$). This is equivalent to demonstrating that the dual variable $\mathbf{v}$ in Problem (3.9) is bounded.

To that end, we will prove that the feasible region of Problem (3.9) is bounded. Note that Problem (3.9) is in an abstract form, and its specific formulation depends on two parameters: k and ε. Without loss of generality (as stated in Remark 1), we will prove that, for any $k = 2^\theta$ with $\theta \in \mathbb{Z}_{++}$, the feasible region of Problem (3.9) is bounded. We will achieve this through mathematical induction, starting with Problem (3.9) when $k = 2$ and then extending the conclusion to the case with a general k. Specifically, when $k = 2$, Problem (3.8) can be written as follows:

$$\max_{\substack{w_1,w_2,\xi^0,\eta^0,\\ \xi^j,\eta^j,\forall j\in[\nu]}} \quad z_1w_1 + z_2w_2 \tag{3.13a}$$

$$(\lambda_1 \geq 0) \quad \text{s.t.} \quad w_1 - \xi^0 \leq 0, \tag{3.13b}$$

$$(\lambda_2 \geq 0) \quad -w_1 - \xi^0 \leq 0, \tag{3.13c}$$

$$(\lambda_3 \geq 0) \quad w_2 - \eta^0 \leq 0, \tag{3.13d}$$

$$(\lambda_4 \geq 0) \quad -w_2 - \eta^0 \leq 0, \tag{3.13e}$$

$$(\omega_j \text{ free}) \quad \cos\left(\frac{\pi}{2^{j+1}}\right)\xi^{j-1} + \sin\left(\frac{\pi}{2^{j+1}}\right)\eta^{j-1} - \xi^j = 0,\ \forall\, j \in [\nu], \tag{3.13f}$$

$$(k_j \geq 0) \quad -\sin\left(\frac{\pi}{2^{j+1}}\right)\xi^{j-1} + \cos\left(\frac{\pi}{2^{j+1}}\right)\eta^{j-1} - \eta^j \leq 0,\ \forall\, j \in [\nu], \tag{3.13g}$$

$$(u_j \geq 0) \quad \sin\left(\frac{\pi}{2^{j+1}}\right)\xi^{j-1} - \cos\left(\frac{\pi}{2^{j+1}}\right)\eta^{j-1} - \eta^j \leq 0,\ \forall\, j \in [\nu], \tag{3.13h}$$

$$(\lambda_5 \geq 0) \quad -\tan\left(\frac{\pi}{2^{\nu+1}}\right)\xi^{\nu} + \eta^{\nu} \leq 0, \tag{3.13i}$$

$$(\lambda_6 \geq 0) \quad \xi^{\nu} \leq 1. \tag{3.13j}$$

We use the notation in the parentheses on the left side of Problem (3.13) to denote the corresponding dual variables for the constraints. Accordingly, we derive the following dual formulation of (3.13), corresponding to Problem (3.9) with $k = 2$.

$$\min_{\substack{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6,\\ k_j,u_j,\omega_j,\forall j\in[\nu]}} \quad \lambda_6 \tag{3.14a}$$

$$\text{s.t.} \quad \lambda_1 - \lambda_2 = z_1, \tag{3.14b}$$

$$\lambda_3 - \lambda_4 = z_2, \tag{3.14c}$$

$$-\lambda_1 - \lambda_2 + \omega_1\cos\left(\frac{\pi}{4}\right) - k_1\sin\left(\frac{\pi}{4}\right) + u_1\sin\left(\frac{\pi}{4}\right) = 0, \tag{3.14d}$$

$$-\lambda_3 - \lambda_4 + \omega_1\sin\left(\frac{\pi}{4}\right) + k_1\cos\left(\frac{\pi}{4}\right) - u_1\cos\left(\frac{\pi}{4}\right) = 0, \tag{3.14e}$$

$$-\omega_j + \omega_{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right) - k_{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right) + u_{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right) = 0, \quad \forall\, j \in [\nu-1], \tag{3.14f}$$

$$-k_j - u_j + \omega_{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right) + k_{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right) - u_{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right) = 0, \quad \forall\, j \in [\nu-1], \tag{3.14g}$$

$$-\omega_\nu - \tan\left(\frac{\pi}{2^{\nu+1}}\right)\lambda_5 + \lambda_6 = 0, \tag{3.14h}$$

$$-k_\nu - u_\nu + \lambda_5 = 0, \tag{3.14i}$$

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6 \geq 0, \tag{3.14j}$$

$$k_j, u_j \geq 0, \ \forall\, j \in [\nu], \tag{3.14k}$$

$$\lambda_6 \leq M. \tag{3.14l}$$

Note that the constraint (3.14l) is redundant but useful to prove the boundedness conclusion. The specific value of the M in (3.14l) can be set by Remark 3. We then have the following lemma holds.

Lemma 4. *For any $M > 0$, the feasible region of Problem* (3.14) *is bounded.*

Proof. By (3.14d) + (3.14e), we have

$$\omega_1\left(\cos\left(\frac{\pi}{4}\right) + \sin\left(\frac{\pi}{4}\right)\right) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4. \tag{3.15}$$

As $\lambda_i \geq 0$ for any $i \in [4]$, we have

$$\omega_1 \geq 0. \tag{3.16}$$

For any $j \in [\nu - 1]$, by (3.14f) / $\tan(\pi/2^{j+2})$ + (3.14g), we have

$$-\frac{\omega_j}{\tan\left(\frac{\pi}{2^{j+2}}\right)} - k_j - u_j + \frac{\omega_{j+1}}{\sin\left(\frac{\pi}{2^{j+2}}\right)} = 0.$$

That is,

$$\omega_{j+1} = \sin\left(\frac{\pi}{2^{j+2}}\right)\left(\frac{\omega_j}{\tan\left(\frac{\pi}{2^{j+2}}\right)} + k_j + u_j\right), \ \forall\, j \in [\nu - 1]. \tag{3.17}$$

By (3.14k), (3.16), and (3.17), we have

$$\omega_j \geq 0, \;\forall\, j \in [\nu].$$

By (3.14h), we have

$$\lambda_5 = \frac{\lambda_6 - \omega_\nu}{\tan\left(\frac{\pi}{2^{\nu+1}}\right)} \leq \frac{\lambda_6}{\tan\left(\frac{\pi}{2^{\nu+1}}\right)} \leq \frac{M}{\tan\left(\frac{\pi}{2^{\nu+1}}\right)} \tag{3.18}$$

and

$$M \geq \lambda_6 = \omega_\nu + \lambda_5 \tan\left(\frac{\pi}{2^{\nu+1}}\right) \geq \omega_\nu. \tag{3.19}$$

By (3.17), we have

$$\omega_j = \tan\left(\frac{\pi}{2^{j+2}}\right)\left(\frac{\omega_{j+1}}{\sin\left(\frac{\pi}{2^{j+2}}\right)} - k_j - u_j\right) \leq \frac{\omega_{j+1}}{\cos\left(\frac{\pi}{2^{j+2}}\right)}, \;\forall\, j \in [\nu-1]. \tag{3.20}$$

By (3.19) and (3.20), we have

$$\omega_j \leq \frac{\omega_\nu}{\prod_{k=j+2}^{\nu+1} \cos\left(\frac{\pi}{2^k}\right)} \leq \frac{M}{\prod_{k=j+2}^{\nu+1} \cos\left(\frac{\pi}{2^k}\right)}, \;\forall\, j \in [\nu-1]. \tag{3.21}$$

By (3.14j), (3.15), and (3.21), we have

$$\lambda_i \leq \sqrt{2}\omega_1 \leq \frac{\sqrt{2}M}{\prod_{k=3}^{\nu+1} \cos\left(\frac{\pi}{2^k}\right)}, \;\forall\, i \in [4].$$

By (3.17), we have

$$k_j = \frac{\omega_{j+1}}{\sin\left(\frac{\pi}{2^{j+2}}\right)} - \frac{\omega_j}{\tan\left(\frac{\pi}{2^{j+2}}\right)} - u_j \leq \frac{\omega_{j+1}}{\sin\left(\frac{\pi}{2^{j+2}}\right)}, \;\forall\, j \in [\nu-1]. \tag{3.22}$$

By (3.21) and (3.22), we have

$$k_j \leq \frac{M}{\sin\left(\frac{\pi}{2^{j+2}}\right) \prod_{k=j+3}^{\nu+1} \cos\left(\frac{\pi}{2^k}\right)}, \ \forall\, j \in [\nu - 2]; \tag{3.23}$$

$$k_j \leq \frac{M}{\sin\left(\frac{\pi}{2^{j+2}}\right)}, \ \forall\, j = \nu - 1. \tag{3.24}$$

Similarly, we also have

$$u_j \leq \frac{M}{\sin\left(\frac{\pi}{2^{j+2}}\right) \prod_{k=j+3}^{\nu+1} \cos\left(\frac{\pi}{2^k}\right)}, \ \forall\, j \in [\nu - 2]; \tag{3.25}$$

$$u_j \leq \frac{M}{\sin\left(\frac{\pi}{2^{j+2}}\right)}, \ \forall\, j = \nu - 1. \tag{3.26}$$

By (3.14i) and (3.18), we have

$$k_\nu \leq \lambda_5 \leq \frac{M}{\tan\left(\frac{\pi}{2^{\nu+1}}\right)}$$

and

$$u_\nu \leq \lambda_5 \leq \frac{M}{\tan\left(\frac{\pi}{2^{\nu+1}}\right)}.$$

Therefore, all the variables in (3.14) are bounded. $\square$

With Lemma 4, we are now ready to prove the following theorem:

Theorem 8. *For any $M > 0$, the feasible region of Problem (3.9) is bounded.*

Proof. We prove this theorem by using mathematical induction. We only give a brief proof here. Detailed proof can be found in Appendix B.4. Recall that Problem (3.9) has the following specific form:

$$
\begin{aligned}
\min \quad & \lambda_{\theta,1}^{6} && \text{(3.27a)}\\
\text{s.t.} \quad & \lambda_{1,i}^{1}-\lambda_{1,i}^{2}=z_{2i-1},\ \forall\, i\in[2^{\theta-1}], && \text{(3.27b)}\\
& \lambda_{1,i}^{3}-\lambda_{1,i}^{4}=z_{2i},\ \forall\, i\in[2^{\theta-1}], && \text{(3.27c)}\\
& \lambda_{\ell,i}^{1}-\lambda_{\ell,i}^{2}-\lambda_{\ell-1,2i-1}^{6}=0,\ \forall\, \ell\in[\theta]\setminus\{1\}, i\in[2^{\theta-\ell}], && \text{(3.27d)}\\
& \lambda_{\ell,i}^{3}-\lambda_{\ell,i}^{4}-\lambda_{\ell-1,2i}^{6}=0,\ \forall\, \ell\in[\theta]\setminus\{1\}, i\in[2^{\theta-\ell}], && \text{(3.27e)}\\
& -\lambda_{\ell,i}^{1}-\lambda_{\ell,i}^{2}+\omega_{\ell,i}^{1}\cos\left(\frac{\pi}{4}\right)-k_{\ell,i}^{1}\sin\left(\frac{\pi}{4}\right)+u_{\ell,i}^{1}\sin\left(\frac{\pi}{4}\right)=0, &&\\
& \qquad\qquad\qquad\qquad \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], && \text{(3.27f)}\\
& -\lambda_{\ell,i}^{3}-\lambda_{\ell,i}^{4}+\omega_{\ell,i}^{1}\sin\left(\frac{\pi}{4}\right)+k_{\ell,i}^{1}\cos\left(\frac{\pi}{4}\right)-u_{\ell,i}^{1}\cos\left(\frac{\pi}{4}\right)=0, &&\\
& \qquad\qquad\qquad\qquad \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], && \text{(3.27g)}\\
& -\omega_{\ell,i}^{j}+\omega_{\ell,i}^{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right)-k_{\ell,i}^{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right)+u_{\ell,i}^{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right)=0, &&\\
& \qquad\qquad\qquad\qquad \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], j\in[\nu-1], && \text{(3.27h)}\\
& -k_{\ell,i}^{j}-u_{\ell,i}^{j}+\omega_{\ell,i}^{j+1}\sin\left(\frac{\pi}{2^{j+2}}\right)+k_{\ell,i}^{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right)-u_{\ell,i}^{j+1}\cos\left(\frac{\pi}{2^{j+2}}\right)=0, &&\\
& \qquad\qquad\qquad\qquad \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], j\in[\nu-1], && \text{(3.27i)}\\
& -\omega_{\ell,i}^{\nu}-\tan\left(\frac{\pi}{2^{\nu+1}}\right)\lambda_{\ell,i}^{5}+\lambda_{\ell,i}^{6}=0,\ \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], && \text{(3.27j)}\\
& -k_{\ell,i}^{\nu}-u_{\ell,i}^{\nu}+\lambda_{\ell,i}^{5}=0,\ \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], && \text{(3.27k)}\\
& \lambda_{\ell,i}^{1},\lambda_{\ell,i}^{2},\lambda_{\ell,i}^{3},\lambda_{\ell,i}^{4},\lambda_{\ell,i}^{5},\lambda_{\ell,i}^{6}\geq 0,\ \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], && \text{(3.27l)}\\
& k_{\ell,i}^{j},u_{\ell,i}^{j}\geq 0,\ \forall\, \ell\in[\theta], i\in[2^{\theta-\ell}], j\in[\nu], && \text{(3.27m)}\\
& \lambda_{\theta,1}^{6}\leq M. && \text{(3.27n)}
\end{aligned}
$$

With $\lambda_{\theta,1}^{6}$ bounded, by Lemma 4, $\lambda_{\theta,1}^{i}$ ($\forall i\in[6]$), $k_{\theta,1}^{j}$ ($\forall j\in[\nu]$), $u_{\theta,1}^{j}$ ($\forall j\in[\nu]$), and $\omega_{\theta,1}^{j}$ ($\forall j\in[\nu]$) are bounded. It follows that $\lambda_{\theta-1,1}^{6}$ and $\lambda_{\theta-1,2}^{6}$ are bounded by (3.27d) and (3.27e). With $\lambda_{\theta-1,1}^{6}$ bounded, by Lemma 4, $\lambda_{\theta-1,1}^{i}$ ($\forall i\in[6]$), $k_{\theta-1,1}^{j}$ ($\forall j\in[\nu]$), $u_{\theta-1,1}^{j}$ ($\forall j\in[\nu]$), and $\omega_{\theta-1,1}^{j}$ ($\forall j\in[\nu]$) are bounded. With $\lambda_{\theta-1,2}^{6}$ bounded, by Lemma 4, we have that $\lambda_{\theta-1,2}^{i}$ ($\forall i\in[6]$), $k_{\theta-1,2}^{j}$ ($\forall j\in[\nu]$), $u_{\theta-1,2}^{j}$ ($\forall j\in[\nu]$), and $\omega_{\theta-1,2}^{j}$ ($\forall j\in[\nu]$)

are bounded. By (3.27d) and (3.27e), we further have $\lambda^6_{\theta-2,1}, \lambda^6_{\theta-2,2}, \lambda^6_{\theta-2,3}$, and $\lambda^6_{\theta-2,4}$ are bounded. It further follows that, by Lemma 4, for all $\tau \in [4]$, $\lambda^i_{\theta-2,\tau}$ ($\forall i \in [6]$), $k^j_{\theta-2,\tau}$ ($\forall j \in [\nu]$), $u^j_{\theta-2,\tau}$ ($\forall j \in [\nu]$), and $\omega^j_{\theta-2,\tau}$ ($\forall j \in [\nu]$) are bounded. Repeating the above process (see details in Appendix B.4), we can conclude that all the variables are bounded. $\square$

Note that M in (3.27n) can be set as $(1+\varepsilon)U_z$ (see Remark 3). Theorem 8 implies that $\mathbf{v}_i$ ($\forall i \in [m]$) in Problem (3.11) are also bounded from above. For each $i \in [m]$, we let K_i be the dimension of $\mathbf{v}_i$. Now we can introduce additional binary variables $\boldsymbol{\beta}_i \in \mathbb{R}^{K_i}$ ($\forall i \in [m]$) such that constraints $\mathbf{v}_i^\top(\mathbf{P}_\varepsilon^{k_i^-+1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^-+1}\mathbf{s}_i + \mathbf{p}_\varepsilon^{k_i^-+1}) = 0$ ($\forall\, i \in [m]$) are equivalent to

$$\mathbf{P}_\varepsilon^{k_i^-+1}\mathbf{w}_i + \mathbf{H}_\varepsilon^{k_i^-+1}\mathbf{s}_i + \mathbf{p}_\varepsilon^{k_i^-+1} \le M\boldsymbol{\beta}_i,\ \forall i \in [m], \tag{3.28}$$

$$\mathbf{v}_i \le M(\mathbf{e} - \boldsymbol{\beta}_i),\ \forall i \in [m], \tag{3.29}$$

$$\beta_{ij} \in \{0,1\},\ \forall i \in [m], j \in [K_i], \tag{3.30}$$

where $M > 0$ is a large enough number. In (3.28), we can set $M = (1+\varepsilon)|\mathbf{P}_\varepsilon^{k_i^-+1}|\mathbf{e} + (1+\varepsilon)|\mathbf{H}_\varepsilon^{k_i^-+1}|\mathbf{e} + |\mathbf{p}_\varepsilon^{k_i^-+1}|$. In (3.29), we can set M as the upper bound of variable $\mathbf{v}_i$ for each $i \in [m]$ by Theorem 8. To obtain such an upper bound, we can first change the objective of the LP problem (3.9), i.e., Problem (3.27), to $\sum_{j=1}^{K_i} v_{ij}$ and obtain the corresponding optimal value ϕ. Then, we can set M in constraints (3.29) as ϕ. Finally, the LPCC (3.11) can be reformulated as the following MILP:

$$\Theta_\varepsilon = \min_{\substack{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u},\\ \mathbf{v},\mathbf{w},\mathbf{s},\mathbf{t},\boldsymbol{\beta}}} \quad \mathbf{c}_0^\top\mathbf{x} \tag{3.31}$$

$$\text{s.t.}\quad (3.11\text{b}) - (3.11\text{h}), (3.3\text{d}) - (3.3\text{e}), (3.28) - (3.30).$$

3.3 An Alternative Approximation

In this section, we present an alternative MILP approximation for Problem (3.1). This alternative MILP approximation typically introduces fewer variables than the first MILP approximation in Section 3.2, although it may result in more constraints when the approximation quality parameter ε is small. The trade-off between the number of variables and the number of constraints is crucial in the context of extended formulations for approximations (Vielma et al. 2017). Thus, we aim to evaluate which MILP approximation performs better in approximating the non-convex QCP (3.1).

We first provide alternative approximations for constraints (3.3b) and (3.3c) while maintaining the same tightness as the original approximations proposed in Sections 3.2.2 and 3.2.3. Next, we introduce an alternative LPCC approximation for Problem (3.1) and demonstrate that it shares the same properties as the previous one in Section 3.2.4, ensuring that all the conclusions from Section 3.2.4 remain applicable here. Finally, we reformulate this alternative LPCC into an MILP.

3.3.1 Polyhedral Approximation of $\|\mathbf{y}\| \leq t$

Note that the first polyhedral approximation of $\|\mathbf{y}\| \leq t$ proposed in Section 3.2.2 is an approximation in a lifted space with an aim to introduce fewer constraints when approximating $\|\mathbf{y}\| \leq t$. Here, we provide an alternative polyhedral approximation of $\|\mathbf{y}\| \leq t$, which focuses on the original space of $(\mathbf{y}, t)$ and aims to introduce fewer variables when approximating $\|\mathbf{y}\| \leq t$. The basic idea is to select a regular m-sided polygon as the polyhedral outer approximation of a 2-dimensional disk. We eventually obtain the following polyhedral approximation of the set L^k (see detailed derivation procedure in

Appendix B.3):

$$\Pi^k(\nu_1,\ldots,\nu_\theta) := \left\{ (\mathbf{y},t) \;\middle|\; \begin{array}{l} \xi_j \geq |y^0_{2j-1}|, \eta_j \geq |y^0_{2j}|, \xi_j \boldsymbol{\alpha}(\nu_1) + \eta_j \boldsymbol{\beta}(\nu_1) \leq y^1_j \mathbf{e},\ \forall j \in \left[2^{\theta-1}\right], \\ y^{\ell-1}_{2j-1}\boldsymbol{\alpha}(\nu_\ell) + y^{\ell-1}_{2j}\boldsymbol{\beta}(\nu_\ell) \leq y^\ell_j \mathbf{e},\ \forall j \in \left[2^{\theta-\ell}\right],\ \ell \in \{2,\ldots,\theta\} \\ \text{for some } \xi_j, \eta_j,\ \forall j \in [2^{\theta-1}],\ y^0_{k+1},\ldots,y^0_{2^\theta}, \\ \hfill \text{and } y^\ell_j,\ \forall j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta-1] \end{array} \right\},$$

where θ is the smallest integer such that $k \leq 2^\theta$, $y^0_i = y_i$ for all $i \in [k]$, $y^\theta_1 = t$, $\boldsymbol{\alpha}(\nu_\ell) := (\cos(\tau\pi/2^{\nu_\ell}))^{2^{\nu_\ell-1}}_{\tau=0} \in \mathbb{R}^{2^{\nu_\ell-1}+1}_+$, and $\boldsymbol{\beta}(\nu_\ell) := (\sin(\tau\pi/2^{\nu_\ell}))^{2^{\nu_\ell-1}}_{\tau=0} \in \mathbb{R}^{2^{\nu_\ell-1}+1}_+$. Note that ν_ℓ ($\forall \ell \in \theta$) are parameters used to control the approximation quality ε_ℓ in level ℓ.

It is important to note that the polyhedral approximation $\Pi^k(\nu_1,\ldots,\nu_\theta)$ has the same approximation quality as $\mathcal{P}^k_\varepsilon$ proposed in Section 3.2.2. To that end, we first show that $\Pi^2(\nu)$ and $\mathcal{P}^2_\varepsilon$ have the same approximation quality when approximating L^2, as stated in the following proposition.

Proposition 10. *The set $\Pi^2(\nu)$ is a polyhedral $\delta(\nu)$-approximation of L^2 with*

$$\delta(\nu) = \frac{1}{\cos\left(\frac{\pi}{2^{\nu+1}}\right)} - 1.$$

Proof. First, we prove that $L^2 \subseteq \Pi^2(\nu)$. For any $(y_1,y_2,y_3) \in L^2$, we set $\xi_0 = |y_1|$ and $\eta_0 = |y_2|$, which satisfy constraints (B.1a)–(B.1b). By the Cauchy-Schwarz inequality, we have

$$\cos\left(\frac{\tau\pi}{2^\nu}\right)\xi_0 + \sin\left(\frac{\tau\pi}{2^\nu}\right)\eta_0 \leq \sqrt{y_1^2+y_2^2}\sqrt{\cos\left(\frac{\tau\pi}{2^\nu}\right)^2 + \sin\left(\frac{\tau\pi}{2^\nu}\right)^2} \leq y_3,\ \forall \tau \in \{0\} \cup [2^{\nu-1}],$$

where the last inequality holds because $(y_1,y_2,y_3) \in L^2$. Thus, (B.1c) is also satisfied and accordingly $L^2 \subseteq \Pi^2(\nu)$.

Next, we show that $(y_1,y_2,y_3) \in \Pi^2(\nu)$ implies that $\sqrt{y_1^2+y_2^2} \leq (1+\delta(\nu))y_3$. By constraints (B.1), we have

$$\cos\left(\frac{\tau\pi}{2^\nu}\right)|y_1| + \sin\left(\frac{\tau\pi}{2^\nu}\right)|y_2| \le y_3,\ \forall \tau \in \{0\}\cup[2^{\nu-1}]. \tag{3.32}$$

Let $\theta_0 := \arccos\left(|y_1|/\sqrt{y_1^2+y_2^2}\right)$, then $|y_1| = \cos(\theta_0)\sqrt{y_1^2+y_2^2}$ and $|y_2| = \sin(\theta_0)\sqrt{y_1^2+y_2^2}$. Therefore,

$$y_3 \ge \left(\cos\left(\frac{\tau\pi}{2^\nu}\right)\cos(\theta_0) + \sin\left(\frac{\tau\pi}{2^\nu}\right)\sin(\theta_0)\right)\sqrt{y_1^2+y_2^2} = \cos\left(\frac{\tau\pi}{2^\nu} - \theta_0\right)\sqrt{y_1^2+y_2^2}$$

for all $\tau \in \{0\}\cup[2^{\nu-1}]$. Since there exists $\tau_0 \in [2^{\nu-1}]$ such that $(\tau_0-1)\pi/2^\nu \le \theta_0 \le \tau_0\pi/2^\nu$, it follows that

$$\begin{aligned}
y_3 &\ge \max\left\{\cos\left(\theta_0 - \frac{(\tau_0-1)\pi}{2^\nu}\right), \cos\left(\frac{\tau_0\pi}{2^\nu} - \theta_0\right)\right\}\sqrt{y_1^2+y_2^2},\\
&= \cos\left(\min\left\{\theta_0 - \frac{(\tau_0-1)\pi}{2^\nu}, \frac{\tau_0\pi}{2^\nu} - \theta_0\right\}\right)\sqrt{y_1^2+y_2^2},\\
&\ge \cos\left(\frac{\pi}{2^{\nu+1}}\right)\sqrt{y_1^2+y_2^2},
\end{aligned}$$

where the last inequality holds because $(\tau_0-1)\pi/2^\nu \le \theta_0 \le \tau_0\pi/2^\nu$ induces $\min\{\theta_0 - (\tau_0-1)\pi/2^\nu, \tau_0\pi/2^\nu - \theta_0\} \le \pi/2^{\nu+1}$. Therefore,

$$\sqrt{y_1^2+y_2^2} \le \frac{y_3}{\cos\left(\frac{\pi}{2^{\nu+1}}\right)} = (1+\delta(\nu))y_3,$$

which completes the proof. □

By Proposition 2.1 of Ben-Tal and Nemirovski (2001) and Proposition 10 above, $\Pi^2(\nu)$ and $\mathcal{P}_\varepsilon^2$ have the same approximation quality when approximating L^2. Because $\Pi^k(\nu_1,\ldots,\nu_\theta)$ shares the same tower-of-variables construction as $\mathcal{P}_\varepsilon^k$, we have $L^k \subseteq \Pi^k(\nu_1,\ldots,\nu_\theta)$, and the overall quality ε is equal to

$$\varepsilon = \prod_{\ell=1}^{\theta}(1+\varepsilon_\ell) - 1 = \prod_{\ell=1}^{\theta}(1+\delta(\nu_\ell)) - 1 = \prod_{\ell=1}^{\theta}\frac{1}{\cos\left(\frac{\pi}{2^{\nu_\ell+1}}\right)} - 1.$$

Comparing it with (3.6) shows that $\Pi^k(\nu_1, \ldots, \nu_\theta)$ has the same approximation quality as $\mathcal{P}^k_\varepsilon$.

Remark 5. *We introduce two additional variables and $2^{\nu-1}+3$ linear constraints in $\Pi^2(\nu)$ to approximate the set L^2. In contrast, $\mathcal{P}^2_\varepsilon$ has $2\nu+2$ additional variables and $3v+6$ linear constraints for its polyhedral approximation of L^2. Clearly, when we choose $\nu \le 5$, $\Pi^2(\nu)$ has fewer variables and constraints compared to $\mathcal{P}^2_\varepsilon$. Notably, when $\nu = 5$, the approximation quality is $\varepsilon = 0.00121$, indicating that for a target approximation quality $\varepsilon \ge 0.00121$, the size of $\Pi^2(\nu)$ is smaller than that of $\mathcal{P}^2_\varepsilon$. For a general k, when the target approximation quality $\varepsilon \ge 0.01$, the number of variables and constraints in the polyhedral approximation $\Pi^k(\nu_1, \ldots, \nu_\theta)$ may be less than that in $\mathcal{P}^k_\varepsilon$ (see details in Table 3.1 in the coming Section 3.4.1). This suggests that the second polyhedral approximation may outperform the first when $\varepsilon \ge 0.01$.*

3.3.2 Approximation of $\|\mathbf{z}\| \ge t$

Recall that $\bar{L}^k := \{(\mathbf{z}, t) \in \mathbb{R}^k \times \mathbb{R} \mid \|\mathbf{z}\| \ge t\}$ is the complement of the $(k+1)$-dimensional SOC. By the definition of the Euclidean norm, we have $\|\mathbf{z}\| = \max_{\mathbf{w}}\{\mathbf{w}^\top \mathbf{z} \mid \|\mathbf{w}\| \le 1\}$. We use the polyhedral approximation given in Section 3.3.1 to represent $\|\mathbf{w}\| \le 1$. As a result, we can provide an approximation of $\|\mathbf{z}\|$ by considering the optimal value of the following problem:

$$
\begin{aligned}
&& \max_{\mathbf{w}, w^0_{k+1}, \ldots, w^0_{2^\theta}, \mathbf{s}} \quad & \mathbf{w}^\top \mathbf{z} && (3.33a)\\
&(u^1_j) & \text{s.t.} \quad & -s^0_{2j-1} + w^0_{2j-1} \le 0,\ \forall j \in \left[2^{\theta-1}\right], && (3.33b)\\
&(u^2_j) & & -s^0_{2j-1} - w^0_{2j-1} \le 0,\ \forall j \in \left[2^{\theta-1}\right], && (3.33c)\\
&(u^3_j) & & -s^0_{2j} + w^0_{2j} \le 0,\ \forall j \in \left[2^{\theta-1}\right], && (3.33d)\\
&(u^4_j) & & -s^0_{2j} - w^0_{2j} \le 0,\ \forall j \in \left[2^{\theta-1}\right], && (3.33e)\\
&(\mathbf{v}^\ell_j) & & s^{\ell-1}_{2j-1}\boldsymbol{\alpha}(\nu_\ell) + s^{\ell-1}_{2j}\boldsymbol{\beta}(\nu_\ell) \le s^\ell_j \mathbf{e},\ \forall\, j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta-1], && (3.33f)
\end{aligned}
$$

$$
\begin{aligned}
&(\mathbf{v}_1^\theta) && s_1^{\theta-1}\boldsymbol{\alpha}(\nu_\theta)+s_2^{\theta-1}\boldsymbol{\beta}(\nu_\theta)\le\mathbf{e}, && (3.33g)\\
& && s_j^\ell\ge 0,\ \forall j\in\left[2^{\theta-\ell}\right],\ \ell\in\{0,\dots,\theta-1\}, && (3.33h)
\end{aligned}
$$

where the symbols in the brackets on the left side are the corresponding dual variables. By the strong duality of linear programming, Problem (3.33) is equivalent to

$$
\begin{aligned}
& & \min_{\mathbf{u},\mathbf{v}} \ & \mathbf{e}^\top\mathbf{v}_1^\theta && (3.34a)\\
&(w_{2j-1}^0) & \text{s.t.}\ & u_j^1-u_j^2=z_{2j-1},\ \forall j\in[2^{\theta-1}], && (3.34b)\\
&(w_{2j}^0) & & u_j^3-u_j^4=z_{2j},\ \forall j\in[2^{\theta-1}], && (3.34c)\\
&(s_j^0) & & -u^1_{\frac{j+1}{2}}-u^2_{\frac{j+1}{2}}+\boldsymbol{\alpha}(\nu_1)^\top\mathbf{v}^1_{\frac{j+1}{2}}\ge 0,\ \forall j\in\{1,3,\dots,2^\theta-1\}, && (3.34d)\\
&(s_j^0) & & -u^3_{\frac{j}{2}}-u^4_{\frac{j}{2}}+\boldsymbol{\beta}(\nu_1)^\top\mathbf{v}^1_{\frac{j}{2}}\ge 0,\ \forall j\in\{2,4,\dots,2^\theta\}, && (3.34e)\\
&(s_j^\ell) & & -\mathbf{e}^\top\mathbf{v}_j^\ell+\boldsymbol{\alpha}(\nu_{\ell+1})^\top\mathbf{v}^{\ell+1}_{\frac{j+1}{2}}\ge 0,\ \forall j\in\left\{1,3,\dots,2^{\theta-\ell}-1\right\},\ell\in[\theta-1], && (3.34f)\\
&(s_j^\ell) & & -\mathbf{e}^\top\mathbf{v}_j^\ell+\boldsymbol{\beta}(\nu_{\ell+1})^\top\mathbf{v}^{\ell+1}_{\frac{j}{2}}\ge 0,\ \forall j\in\left\{2,4,\dots,2^{\theta-\ell}\right\},\ell\in[\theta-1], && (3.34g)\\
& & & u_j^1,u_j^2,u_j^3,u_j^4\ge 0,\ \forall j\in[2^{\theta-1}],\ \mathbf{v}_j^\ell\ge 0,\ \forall j\in\left[2^{\theta-\ell}\right],\ell\in[\theta], && (3.34h)\\
& & & \mathbf{e}^\top\mathbf{v}_1^\theta\le M, && (3.34i)
\end{aligned}
$$

where $M>0$ is a large enough number and we can set $M=(1+\varepsilon)U_z$ in (3.34i) (see Remark 3). Here (3.34i) is a redundant constraint to bound the objective function (3.34). Note that the primal problem (3.33) is equivalent to the dual problem (3.34) by the strong duality of linear programming. However, as the feasible region of the primal problem (3.33) is bounded and the feasible region of the dual problem (3.34) can be unbounded, we add the redundant constraint (3.34i) to ensure that the latter is also bounded, as shown in the following proposition.

Proposition 11. *The feasible region of Problem (3.34) is bounded.*

Proof. First, constraint (3.34i) shows that $\mathbf{v}_1^\theta$ is upper-bounded. As $\boldsymbol{\alpha}(\nu_\theta)$ and $\boldsymbol{\beta}(\nu_\theta)$ are nonnegative, we then have that $\mathbf{v}_1^{\theta-1}$ and $\mathbf{v}_2^{\theta-1}$ are upper-bounded by constraints (3.34f) and (3.34g). Repeating applying constraints (3.34f) and (3.34g), we have $\mathbf{v}_j^\ell$ is upper-bounded for any $j \in [2^{\theta-\ell}]$ and $\ell \in [\theta]$. Second, constraints (3.34h) show that $\mathbf{v}_j^\ell$ is lower-bounded for any $j \in [2^{\theta-\ell}]$ and $\ell \in [\theta]$. Therefore, $\mathbf{v}_j^\ell$ $(\forall j \in [2^{\theta-\ell}], \ell \in [\theta])$ are bounded. In the end, by constraints (3.34d), (3.34e), and (3.34h), we have that $u_j^1, u_j^2, u_j^3, u_j^4$ $(\forall j \in [2^{\theta-1}])$ are also bounded, which completes the proof. □

Finally, by applying the above approximations in Sections 3.3.1 and 3.3.2, we have an alternative approximation of Problem (3.1) as follows:

$$
\begin{aligned}
\min_{\substack{\mathbf{x},\mathbf{y},\mathbf{z},\\ \boldsymbol{\xi},\boldsymbol{\eta},\mathbf{u},\mathbf{v},\mathbf{s}}} \quad & \mathbf{c}_0^\top \mathbf{x} && \text{(3.35a)}\\
\text{s.t.} \quad & \xi_{i,j} \geq |y_{i,2j-1}^0|,\ \eta_{i,j} \geq |y_{i,2j}^0|,\ \xi_{i,j}\boldsymbol{\alpha}(\nu_1) + \eta_{i,j}\boldsymbol{\beta}(\nu_1) \leq y_{i,j}^1\mathbf{e}, \\
& \qquad \forall j \in \left[2^{\theta_i^y-1}\right],\ i \in [m], && \text{(3.35b)}\\
& y_{i,2j-1}^{\ell-1}\boldsymbol{\alpha}(\nu_\ell) + y_{i,2j}^{\ell-1}\boldsymbol{\beta}(\nu_\ell) \leq y_{i,j}^\ell \mathbf{e}, \\
& \qquad \forall j \in \left[2^{\theta_i^y-\ell}\right],\ \ell \in \{2,\dots,\theta_i^y\},\ i \in [m], && \text{(3.35c)}\\
& -s_{i,2j-1}^0 + w_{i,2j-1}^0 \leq 0, -s_{i,2j-1}^0 - w_{i,2j-1}^0 \leq 0, \forall j \in \left[2^{\theta_i^z-1}\right], i \in [m], && \text{(3.35d)}\\
& -s_{i,2j}^0 + w_{i,2j}^0 \leq 0, -s_{i,2j}^0 - w_{i,2j}^0 \leq 0,\ \forall j \in \left[2^{\theta_i^z-1}\right],\ i \in [m], && \text{(3.35e)}\\
& s_{i,2j-1}^{\ell-1}\boldsymbol{\alpha}(\nu_\ell) + s_{i,2j}^{\ell-1}\boldsymbol{\beta}(\nu_\ell) \leq s_{i,j}^\ell \mathbf{e}, \forall j \in \left[2^{\theta_i^z-\ell}\right], \ell \in [\theta_i^z-1], i \in [m], && \text{(3.35f)}\\
& s_{i,1}^{\theta_i^z-1}\boldsymbol{\alpha}(\nu_{\theta_i^z}) + s_{i,2}^{\theta_i^z-1}\boldsymbol{\beta}(\nu_{\theta_i^z}) \leq \mathbf{e}, \forall i \in [m], && \text{(3.35g)}\\
& s_{i,j}^\ell \geq 0, \forall j \in \left[2^{\theta_i^z-\ell}\right],\ \ell \in \{0,\dots,\theta_i^z-1\},\ i \in [m], && \text{(3.35h)}\\
& u_{i,j}^1 - u_{i,j}^2 = z_{i,2j-1},\ u_{i,j}^3 - u_{i,j}^4 = z_{i,2j},\ \forall j \in [2^{\theta_i^z-1}],\ i \in [m], && \text{(3.35i)}\\
& -u_{i,\frac{j+1}{2}}^1 - u_{i,\frac{j+1}{2}}^2 + \boldsymbol{\alpha}(\nu_1)^\top \mathbf{v}_{i,\frac{j+1}{2}}^1 \geq 0, \\
& \qquad \forall j \in \{1,3,\dots,2^{\theta_i^z}-1\},\ i \in [m], && \text{(3.35j)}\\
& -u_{i,\frac{j}{2}}^3 - u_{i,\frac{j}{2}}^4 + \boldsymbol{\beta}(\nu_1)^\top \mathbf{v}_{i,\frac{j}{2}}^1 \geq 0, \forall j \in \{2,4,\dots,2^{\theta_i^z}\}, i \in [m], && \text{(3.35k)}\\
& -\mathbf{e}^\top \mathbf{v}_{i,j}^\ell + \boldsymbol{\alpha}(\nu_{\ell+1})^\top \mathbf{v}_{i,\frac{j+1}{2}}^{\ell+1} \geq 0, \\
& \qquad \forall j \in \left\{1,3,\dots,2^{\theta_i^z-\ell}-1\right\}, \ell \in [\theta_i^z-1],\ i \in [m], && \text{(3.35l)}
\end{aligned}
$$

$$-\mathbf{e}^\top \mathbf{v}_{i,j}^\ell + \boldsymbol{\beta}(\nu_{\ell+1})^\top \mathbf{v}_{i,\frac{j}{2}}^{\ell+1} \geq 0,$$

$$\forall j \in \left\{2, 4, \dots, 2^{\theta_i^z - \ell}\right\}, \ell \in [\theta_i^z - 1], i \in [m], \tag{3.35m}$$

$$u_{i,j}^1, u_{i,j}^2, u_{i,j}^3, u_{i,j}^4 \geq 0,$$

$$\forall j \in [2^{\theta_i^z - 1}],\ \mathbf{v}_{i,j}^\ell \geq 0,\ \forall j \in \left[2^{\theta_i^z - \ell}\right], \ell \in [\theta_i^z],\ i \in [m], \tag{3.35n}$$

$$t_i \leq \mathbf{e}^\top \mathbf{v}_{i,1}^\theta \leq M,\ \forall i \in [m], \tag{3.35o}$$

$$u_{i,j}^1(-s_{i,2j-1}^0 + w_{i,2j-1}^0) = 0, \forall j \in \left[2^{\theta_i^z - 1}\right],\ i \in [m], \tag{3.35p}$$

$$u_{i,j}^2(-s_{i,2j-1}^0 - w_{i,2j-1}^0) = 0,\ \forall j \in \left[2^{\theta_i^z - 1}\right],\ i \in [m], \tag{3.35q}$$

$$u_{i,j}^3(-s_{i,2j}^0 + w_{i,2j}^0) = 0,\ \forall j \in \left[2^{\theta_i^z - 1}\right],\ i \in [m], \tag{3.35r}$$

$$u_{i,j}^4(-s_{i,2j}^0 - w_{i,2j}^0) = 0,\ \forall j \in \left[2^{\theta_i^z - 1}\right],\ i \in [m], \tag{3.35s}$$

$$\mathbf{v}_{i,j}^\ell(s_{i,2j-1}^{\ell-1}\boldsymbol{\alpha}(\nu_\ell) + s_{i,2j}^{\ell-1}\boldsymbol{\beta}(\nu_\ell) - s_{i,j}^\ell \mathbf{e}) = 0, \forall j \in \left[2^{\theta_i^z - \ell}\right],\ \ell \in [\theta_i^z - 1],\ i \in [m], \tag{3.35t}$$

$$\mathbf{v}_{i,1}^{\theta_i^z}(s_{i,1}^{\theta_i^z - 1}\boldsymbol{\alpha}(\nu_{\theta_i^z}) + s_{i,2}^{\theta_i^z - 1}\boldsymbol{\beta}(\nu_{\theta_i^z}) - \mathbf{e}) = 0,\ \forall i \in [m], \tag{3.35u}$$

$$s_j^0(-u_{i,\frac{j+1}{2}}^1 - u_{i,\frac{j+1}{2}}^2 + \boldsymbol{\alpha}(\nu_1)^\top \mathbf{v}_{i,\frac{j+1}{2}}^1) = 0, \forall j \in \{1, 3, \dots, 2^{\theta_i^z} - 1\},\ i \in [m], \tag{3.35v}$$

$$s_{i,j}^0(-u_{i,\frac{j}{2}}^3 - u_{i,\frac{j}{2}}^4 + \boldsymbol{\beta}(\nu_1)^\top \mathbf{v}_{i,\frac{j}{2}}^1) = 0,\ \forall j \in \{2, 4, \dots, 2^{\theta_i^z}\},\ i \in [m], \tag{3.35w}$$

$$s_{i,j}^\ell(-\mathbf{e}^\top \mathbf{v}_{i,j}^\ell + \boldsymbol{\alpha}(\nu_{\ell+1})^\top \mathbf{v}_{i,\frac{j+1}{2}}^{\ell+1}) = 0,$$

$$\forall j \in \left\{1, 3, \dots, 2^{\theta_i^z - \ell} - 1\right\}, \ell \in [\theta_i^z - 1],\ i \in [m], \tag{3.35x}$$

$$s_{i,j}^\ell(-\mathbf{e}^\top \mathbf{v}_{i,j}^\ell + \boldsymbol{\beta}(\nu_{\ell+1})^\top \mathbf{v}_{i,\frac{j}{2}}^{\ell+1}) = 0, \forall j \in \left\{2, 4, \dots, 2^{\theta_i^z - \ell}\right\}, \ell \in [\theta_i^z - 1],\ i \in [m], \tag{3.35y}$$

(3.3d) – (3.3e).

where θ_i^y and θ_i^z are the corresponding θ in $\mathbf{y}_i$ and $\mathbf{z}_i$. Note that the LPCC approximation (3.35) has the same approximation quality as (3.11). Thus, the asymptotic convergence property (Proposition 9 and Theorem 7) also holds for (3.35). By the boundedness result in Proposition 11, we can also obtain an MILP reformulation for (3.35) by the Big-M method.

As we explain in Remark 5, when the target approximation quality is larger than 0.01, the second MILP approximation has fewer variables and constraints than the first one. However, when the target approximation quality is no larger than 0.01, the second MILP approximation results in more constraints while still fewer variables. Depending on a trade-off between the number of variables and the number of constraints, the MILP approximations developed in Sections 3.2 and 3.3 can be both useful and relevant.

3.4 Computational Experiments

In this section, we present computational analyses of our proposed approximations for non-convex QCPs. First, we perform extensive experiments to evaluate the effectiveness of these approximations over various instances. Then, we perform sensitivity analyses to examine the impact of various parameters, including the number of convex quadratic constraints m_1, the number of non-convex quadratic constraints m_2, and the approximation quality parameter ε, on the performance of our approximations. Finally, to demonstrate the applicability of our methods, we apply the proposed approximations to solve two specific problems: the joint decision and estimation problem and the two-trust-region subproblem. All mathematical models are implemented in Python (ver. 3.10.0) using the Gurobi solver (ver. 10.0.3) on a high-performance computing cluster consisting of 13 nodes. Each node has 24 CPU cores and 400 GB of usable memory. The operating system is CentOS 7.6, and the CPU used is an Intel(R) Xeon(R) E5-2699 CPU @ 2.30GHz processor. A time limit of 2 hours is set for each run. All the data instances and source code can be found in Jiang et al. (2025b).

3.4.1 Simplification of Approximation and Selection of ν_ℓ

Recall that we propose two polyhedral approximations for the SOC constraint $\|\mathbf{y}\| \leq t$ in Sections 3.2 and 3.3, respectively. Henceforth, we refer to the first polyhedral approximation in Section 3.2.2 as "P1" and its corresponding LPCC/MILP approximation (i.e., Problem (3.11)/(3.31)) as "A1." Similarly, the second polyhedral approximation in Section 3.3.1 is denoted by "P2," with its corresponding LPCC/MILP approximation (i.e., Problem (3.35)) referred to as "A2." It is important to note that the number of constraints in P1 and P2 significantly affects the computational time because such a number determines the number of complementarity constraints in the LPCCs, which further determines the number of binary variables introduced in the MILPs. Therefore, we aim to reduce the number of constraints in P1 and P2 as much as possible while ensuring that the overall approximation quality does not exceed a predefined threshold, i.e., ε.

One possible approach to simplifying the polyhedral approximation for P1 is detailed below. First, in constraints (3.5), for each ℓ, i, and j, we substitute the variable $\xi^j_{\ell,i}$ and accordingly eliminate (3.5e). Second, given that $y^\ell_i \geq 0$ ($\forall \ell \in [\theta]$), we can drop the second and fourth constraints in (3.5) for all $\ell \in \{2, \ldots, \theta\}$ and $i \in [2^{\theta-\ell}]$. After rearrangement, (3.5) is equivalent to the following system:

$$
\begin{aligned}
&\xi^0_{1,i} \geq y^0_{2i-1}, \ \forall i \in [2^{\theta-1}],\\
&\xi^0_{1,i} \geq -y^0_{2i-1}, \ \forall i \in [2^{\theta-1}],\\
&\eta^0_{1,i} \geq y^0_{2i}, \ \forall i \in [2^{\theta-1}],\\
&\eta^0_{1,i} \geq -y^0_{2i}, \ \forall i \in [2^{\theta-1}],\\
&\xi^0_{\ell,i} \geq \frac{1}{\sin\left(\frac{\pi}{2^{\nu_{\ell-1}+1}}\right)} \left(\frac{1}{2^{\nu_{\ell-1}}}\, \xi^0_{\ell-1,2i-1} + \sum_{k=0}^{\nu_{\ell-1}-1} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{\nu_{\ell-1}-1-k}} \eta^k_{\ell-1,2i-1} \right), \forall i \in [2^{\theta-\ell}], \ell = 2, \ldots, \theta,\\
&\eta^0_{\ell,i} \geq \frac{1}{\sin\left(\frac{\pi}{2^{\nu_{\ell-1}+1}}\right)} \left(\frac{1}{2^{\nu_{\ell-1}}}\, \xi^0_{\ell-1,2i} + \sum_{k=0}^{\nu_{\ell-1}-1} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{\nu_{\ell-1}-1-k}} \eta^k_{\ell-1,2i} \right), \forall i \in [2^{\theta-\ell}], \ell = 2, \ldots, \theta,\\
&\eta^j_{\ell,i} \geq -\frac{1}{2\cos\left(\frac{\pi}{2^{j+1}}\right)} \left(\frac{1}{2^{j-1}}\, \xi^0_{\ell,i} + \sum_{k=0}^{j-2} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{j-2-k}} \eta^k_{\ell,i} \right) + \cos\left(\frac{\pi}{2^{j+1}}\right) \eta^{j-1}_{\ell,i},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall j \in [\nu_\ell], i \in [2^{\theta-\ell}], \ell \in [\theta],\\
&\eta^j_{\ell,i} \geq \frac{1}{2\cos\left(\frac{\pi}{2^{j+1}}\right)} \left(\frac{1}{2^{j-1}}\, \xi^0_{\ell,i} + \sum_{k=0}^{j-2} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{j-2-k}} \eta^k_{\ell,i} \right) - \cos\left(\frac{\pi}{2^{j+1}}\right) \eta^{j-1}_{\ell,i},
\end{aligned}
$$

$$\forall j \in [\nu_\ell], i \in [2^{\theta-\ell}], \ell \in [\theta],$$

$$\eta_{\ell,i}^{\nu_\ell} \leq \frac{1}{\cos\left(\frac{\pi}{2^{\nu_\ell+1}}\right)} \left(\frac{1}{2^{\nu_\ell}}\, \xi_{\ell,i}^0 + \sum_{k=0}^{\nu_\ell-1} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{\nu_\ell-1-k}} \eta_{\ell,i}^k \right), \ \forall i \in [2^{\theta-\ell}], \ell \in [\theta-1],$$

$$\eta_{\theta,1}^{\nu_\theta} \leq \frac{1}{\cos\left(\frac{\pi}{2^{\nu_\theta+1}}\right)} \left(\frac{1}{2^{\nu_\theta}}\, \xi_{\theta,1}^0 + \sum_{k=0}^{\nu_\theta-1} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{\nu_\theta-1-k}} \eta_{\theta,1}^k \right),$$

$$t \geq \frac{1}{\sin\left(\frac{\pi}{2^{\nu_\theta+1}}\right)} \left(\frac{1}{2^{\nu_\theta}}\, \xi_{\theta,1}^0 + \sum_{k=0}^{\nu_\theta-1} \frac{\sin^2\left(\frac{\pi}{2^{k+2}}\right)}{2^{\nu_\theta-1-k}} \eta_{\theta,1}^k \right).$$

We denote the above simplified P1 by "S-P1." Note that P2 does not contain equality constraints, making it difficult to simplify. Thus, the simplification is only applied to P1 in this context.

Another approach to reducing the number of constraints in our approximations involves optimizing the selection of ν_ℓ for any $\ell \in [\theta]$. Specifically, we propose three strategies for selecting ν_ℓ for all $\ell \in [\theta]$, ensuring that the overall approximation quality remains below a predefined threshold ε:

- The "Decision Rule" strategy: We set $\varepsilon_\ell = k^\ell \varepsilon$ ($\forall \ell \in [\theta]$), where ε_ℓ represents the quality of the level ℓ and $0 < k \leq 1$ is a scalar to be optimized. A larger k leads to a larger ε_ℓ, which results in a smaller ν_ℓ and fewer constraints. We choose the largest k such that

$$\prod_{\ell=1}^{\theta} (1 + k^\ell \varepsilon) - 1 \leq \varepsilon. \tag{3.36}$$

  In the end, we set $\nu_\ell = \lceil \log_2(\pi / \arccos(1/(1+\epsilon_\ell))) - 1 \rceil$ to satisfy (3.7).

- The "Same" strategy: We set $\nu_\ell = \nu$ ($\forall \ell \in [\theta]$) such that

$$\left(\frac{1}{\cos\left(\frac{\pi}{2^{\nu+1}}\right)} \right)^\theta - 1 \leq \varepsilon.$$

  Clearly, the smallest choice of ν is $\nu = \lceil \log_2(\pi / \arccos(1/\sqrt[\theta]{1+\epsilon})) - 1 \rceil$.

- The "Primary" strategy: We let $\nu_\ell = \lfloor g\ell \ln \frac{2}{\varepsilon} \rfloor$ $(\forall \ell \in [\theta])$ by following Ben-Tal and Nemirovski (2001), where g is a scalar to be optimized. We choose the smallest g such that

$$\prod_{\ell=1}^{\theta} \frac{1}{\cos\left(\frac{\pi}{2^{\nu_\ell+1}}\right)} - 1 \leq \varepsilon.$$

To summarize, we have three polyhedral approximations (P1, S-P1, and P2) for the SOC constraint $\|\mathbf{y}\| \leq t$ and three strategies to choose ν_ℓ $(\forall \ell \in [\theta])$. We compare their performance in Table 3.1. In this table, K represents the dimension of the SOC constraint $\|\mathbf{y}\| \leq t$ (i.e., the sum of the dimensions of $\mathbf{y}$ and t) and ε represents the target approximation quality parameter ε. The paired values represent the number of variables and the number of constraints in the corresponding polyhedral approximation. From Table 3.1, we have the following observations. First, the "Decision Rule" strategy consistently outperforms the other strategies. For all $K \in \{50, 200, 400\}$ and $\varepsilon \in \{0.1, 0.01, 0.001\}$, the "Decision Rule" strategy introduces fewer variables and constraints for the three polyhedral approximations compared to the "Same" strategy and the "Primary" strategy. Therefore, we will only focus on the "Decision Rule" strategy in the subsequent discussion. Second, S-P1 outperforms P1. For all $K \in \{50, 200, 400\}$ and $\varepsilon \in \{0.1, 0.01, 0.001\}$, both the number of variables and the number of constraints in S-P1 are less than those in P1. Third, P2 consistently has fewer variables than S-P1. However, the number of constraints in P2 increases rapidly. When $\varepsilon = 0.1$, the number of constraints in P2 is about half that of S-P1. When $\varepsilon = 0.01$, P2 and S-P1 have a similar number of constraints. When $\varepsilon = 0.001$, P2 has approximately twice as many constraints as S-P1.

Based on the above observations, *we will utilize the "Decision Rule" strategy to select ν_ℓ and simplify P1 into S-P1 for the first LPCC/MILP approximation A1 in the following experiments.*

Table 3.1: Problem Sizes of Polyhedral Approximations with Different Rules and Parameters

		Decision Rule			Same			Primary		
K	ε	P1	S-P1	P2	P1	S-P1	P2	P1	S-P1	P2
50	0.1	(637, 1146)	(383, 638)	(127, 224)	(757, 1386)	(443, 758)	(127, 504)	(733, 1338)	(431, 734)	(127, 1056)
	0.01	(849, 1570)	(489, 850)	(127, 896)	(883, 1638)	(506, 884)	(127, 1008)	(1213, 2298)	(671, 1214)	(127, 599424)
	0.001	(1033, 1938)	(581, 1034)	(127, 2176)	(1135, 2142)	(632, 1136)	(127, 4032)	(1693, 3258)	(911, 1694)	too large
200	0.1	(2601, 4690)	(1557, 2602)	(511, 1024)	(3061, 5610)	(1787, 3062)	(511, 2040)	(3029, 5546)	(1771, 3030)	(511, 16512)
	0.01	(3453, 6394)	(1983, 3454)	(511, 3968)	(3571, 6630)	(2042, 3572)	(511, 4080)	(5037, 9562)	(2775, 5038)	too large
	0.001	(4209, 7906)	(2361, 4210)	(511, 10752)	(4591, 8670)	(2552, 4592)	(511, 16320)	(7045, 13578)	(3779, 7046)	too large
400	0.1	(5221, 9418)	(3123, 5222)	(1023, 2176)	(6133, 11242)	(3579, 6134)	(1023, 4088)	(6097, 11170)	(3561, 6098)	(1023, 65792)
	0.01	(6927, 12830)	(3976, 6928)	(1023, 8064)	(8177, 15330)	(4601, 8178)	(1023, 16352)	(10149, 19274)	(5587, 10150)	too large
	0.001	(8445, 15866)	(4735, 8446)	(1023, 23552)	(9199, 17374)	(5112, 9200)	(1023, 32704)	(14201, 27378)	(7613, 14202)	too large

3.4.2 Instance Generation and Table Header Description

We now consider a general QCP with m_1 convex and m_2 non-convex quadratic constraints as well as a convex bounded constraint (i.e., $\|\mathbf{x}\| \leq U$), as shown below:

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \mathbf{c}_0^\top \mathbf{x} && \text{(3.37a)}\\
\text{s.t.} \quad & \mathbf{x}^\top \mathbf{Q}_i^{\mathrm{c}} \mathbf{x} + 2{\mathbf{c}_i^{\mathrm{c}}}^\top \mathbf{x} \leq b_i^{\mathrm{c}}, \ \forall i \in [m_1], && \text{(3.37b)}\\
& \mathbf{x}^\top \mathbf{Q}_i^{\mathrm{n}} \mathbf{x} + 2{\mathbf{c}_i^{\mathrm{n}}}^\top \mathbf{x} \leq b_i^{\mathrm{n}}, \ \forall i \in [m_2], && \text{(3.37c)}\\
& \|\mathbf{x}\| \leq U, && \text{(3.37d)}
\end{aligned}
$$

where $\mathbf{Q}_i^{\mathrm{c}}$ is a PSD matrix for any $i \in [m_1]$ and $\mathbf{Q}_i^{\mathrm{n}}$ is an indefinite matrix for any $i \in [m_2]$. We follow a predefined set of rules to generate data (i.e., $\mathbf{Q}_i^{\mathrm{c}}, \mathbf{Q}_i^{\mathrm{n}}, \mathbf{c}_0, \mathbf{c}_i^{\mathrm{c}}, \mathbf{c}_i^{\mathrm{n}}, b_i^{\mathrm{c}}, b_i^{\mathrm{n}}$) for our experiments. To generate the matrices $\mathbf{Q}_i^{\mathrm{c}}$ and $\mathbf{Q}_i^{\mathrm{n}}$, we proceed as follows. Given a dimension N of the decision variable $\mathbf{x}$, our goal is to generate a matrix $\mathbf{Q}$ with $N - N_1$ positive eigenvalues and N_1 negative eigenvalues (i.e., $N_1 = 0$ leads to a PSD matrix). First, we generate two orthogonal matrices $\mathbf{Y}$ and $\mathbf{Z}$ of size $N \times N$. Then, we generate two diagonal matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$. The matrix $\mathbf{\Lambda}_1$ contains $N - N_1$ non-zero values and N_1 zeros along the diagonal, while $\mathbf{\Lambda}_2$ contains $N - N_1$ zeros and N_1 non-zero values along the diagonal. We obtain $\mathbf{A}$ by right multiplying $\mathbf{Y}$ to $\mathbf{\Lambda}_1$, and $\mathbf{B}$ by right multiplying $\mathbf{Z}$ to $\mathbf{\Lambda}_2$. Finally, we set $\mathbf{Q} = \mathbf{A}^\top\mathbf{A} - \mathbf{B}^\top\mathbf{B}$, ensuring that $\mathbf{Q}$ has $N - N_1$ positive eigenvalues and N_1 negative eigenvalues. For vectors $\mathbf{c}_0$, $\mathbf{c}_i^{\mathrm{c}}$ ($\forall i \in [m_1]$), and $\mathbf{c}_i^{\mathrm{n}}$ ($\forall i \in [m_2]$), scalars b_i^{c}

($\forall i \in [m_1]$) and b_i^{n} ($\forall i \in [m_2]$), and diagonal matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$, each element therein is uniformly generated from the interval $[-100, 100]$. We set the value of U to be twice the value of N (i.e., $U = 2 \times N$) for our experiments. In addition to the randomly generated instances above, we also evaluate our MILP approximations on the unitbox instances from Bao et al. (2009) (see details in Appendix B.5).

We set $m_1 = 4$, $m_2 = 2$, and $\varepsilon = 0.01$ for all the experiments here, while sensitivity analyses over these three parameters will be presented in Section 3.4.4. We consider six combinations of parameters N and N_1, with N taking values from $\{100, 150, 200\}$ and N_1 from $\{1, 3\}$. For each combination of N and N_1, we generate ten instances and report the average result. To demonstrate the effectiveness of our proposed approximations, we conduct extensive experiments and compare their performance with the Gurobi solver in solving Problem (3.37). On the one hand, we use the Gurobi solver with default parameters to solve Problem (3.37) directly. On the other hand, we use Gurobi to solve our proposed MILP approximations, which provide lower bounds. Note that each of our MILP approximations can be divided into two parts: the first part includes the LP constraints from the polyhedral approximation of $\|\mathbf{y}\| \leq t$, and the second part includes the MILP constraints from the approximation of $\|\mathbf{z}\| \geq t$. To evaluate the independent performance of each part, we further solve Problem (3.37) in the following two cases: (i) approximating the $\|\mathbf{y}\| \leq t$ part while keeping the $\|\mathbf{z}\| \geq t$ part unchanged and (ii) approximating the $\|\mathbf{z}\| \geq t$ part while keeping the $\|\mathbf{y}\| \leq t$ part unchanged.

To further demonstrate the advantages of our approximations, we also compare our MILP approximations with the standard LP, SOCP, and SDP relaxations of the original Problem (3.1). First, we consider the following standard LP relaxation of Problem (3.1):

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{X}} \quad & \mathbf{c}_0^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{X} \bullet \mathbf{Q}_i + 2\mathbf{c}_i^\top \mathbf{x} \leq b_i, \ \forall\, i \in [m], \\
& X_{ij} - u_j x_i - l_i x_j \leq -u_j l_i, \ \forall i \in [N], j \in [N],
\end{aligned}
$$

$$X_{ij} - l_j x_i - u_i x_j \leq -l_j u_i, \ \forall i \in [N], j \in [N],$$

$$X_{ij} - u_j x_i - u_i x_j \geq -u_j u_i, \ \forall i \in [N], j \in [N],$$

$$X_{ij} - l_j x_i - l_i x_j \geq -l_j l_i, \ \forall i \in [N], j \in [N],$$

where l_i (resp. u_i) denotes the lower (resp. upper) bound of x_i for any $i \in [N]$. For our randomly generated instances, we set $l_i = -2N$ and $u_i = 2N$ by the bounded constraint $\|\mathbf{x}\| \leq 2N$. For the unitbox instances (see Appendix B.5), we set $l_i = 0$ and $u_i = 1$. Second, for the SOCP relaxation, we relax Problem (3.1) into the following convex program (Kim and Kojima 2001):

$$\begin{aligned}
\min_{\mathbf{x},\mathbf{z}} \quad & \mathbf{c}_0^\top \mathbf{x} \\
\text{s.t.} \quad & \mathbf{x}^\top \mathbf{Q}_i^+ \mathbf{x} + \sum_{j=l_i+1}^{n} \lambda_{ij} z_{ij} + 2\mathbf{c}_i^\top \mathbf{x} \leq b_i, \ \forall i \in [m], \\
& \mathbf{x}^\top (\mathbf{u}_{ij} \mathbf{u}_{ij}^\top) \mathbf{x} \leq z_{ij}, \ \forall i \in [m], j \in \{l_i + 1, \dots, n\}, \\
& \sum_{j=l_i+1}^{n} z_{ij} \leq \rho_{max}, \forall i \in [m],
\end{aligned}$$

where for each $i \in [m]$, $\mathbf{Q}_i = \sum_{j=1}^{n} \lambda_{ij} \mathbf{u}_{ij} \mathbf{u}_{ij}^\top$ with λ_{ij} ($\forall j \in [n]$) representing the eigenvalues in non-increasing order and $\mathbf{u}_{ij}$ ($\forall j \in [n]$) representing the corresponding eigenvectors, l_i denotes the number of nonnegative eigenvalues, and $\mathbf{Q}_i^+ = \sum_{j=1}^{l_i} \lambda_{ij} \mathbf{u}_{ij} \mathbf{u}_{ij}^\top$. Note that ρ_{max} should be appropriately selected to bound z_{ij} from above (Kim and Kojima 2001). For our randomly generated instances, we set $\rho_{max} = 2N$. For the unitbox instances (see Appendix B.5), we set $\rho_{max} = N$. Third, for the SDP relaxation, we introduce $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and relax it to the constraint $\mathbf{X} \succeq \mathbf{x}\mathbf{x}^\top$. Because the Gurobi solver is unable to solve SDP problems, we utilize the Mosek solver (ver. 10.1) with default parameters to solve the SDP problem.

To summarize, for each instance described above, we consider the following experiments:

- We use the Gurobi solver with default parameters to solve Problem (3.37) directly.

- We use the Gurobi solver with default parameters to solve the LP relaxation of Problem (3.37).

- We use the Gurobi solver with default parameters to solve the SOCP relaxation of Problem (3.37).

- We use the Mosek solver with default parameters to solve the SDP relaxation of Problem (3.37).

- We reformulate Problem (3.37) into the form of (3.3), where constraints (3.3b) (i.e., $\|\mathbf{y}_i\| \leq t_i$, $\forall i \in [m]$) are approximated using the first polyhedral approximation proposed in Section 3.2.2 and constraints (3.3c) (i.e., $\|\mathbf{z}_i\| \geq t_i$, $\forall i \in [m]$) are not approximated. We then use the Gurobi solver with default parameters to solve this approximation.

- We reformulate Problem (3.37) into the form of (3.3), where constraints (3.3c) (i.e., $\|\mathbf{z}_i\| \geq t_i$, $\forall i \in [m]$) are approximated using MILP constraints based on Sections 3.2.3 and 3.2.5 and constraints (3.3b) (i.e., $\|\mathbf{y}_i\| \leq t_i$, $\forall i \in [m]$) are not approximated. We then use the Gurobi solver with default parameters to solve this approximation.

- We approximate Problem (3.37) with the MILP approximation (3.31) in Section 3.2 (i.e., both constraints (3.3b) and (3.3c) are approximated) and use the Gurobi solver with default parameters to solve this approximation.

- We reformulate Problem (3.37) into the form of (3.3), where constraints (3.3b) (i.e., $\|\mathbf{y}_i\| \leq t_i$, $\forall i \in [m]$) are approximated using the second polyhedral approximation proposed in Section 3.3.1 and constraints (3.3c) (i.e., $\|\mathbf{z}_i\| \geq t_i$, $\forall i \in [m]$) are not approximated. We then use the Gurobi solver with default parameters to solve this approximation.

- We reformulate Problem (3.37) into the form of (3.3), where constraints (3.3c) (i.e., $\|\mathbf{z}_i\| \geq t_i$, $\forall i \in [m]$) are approximated using MILP constraints based on Section

3.3.2 and constraints (3.3b) (i.e., $\|\mathbf{y}_i\| \leq t_i$, $\forall i \in [m]$) are not approximated. We then use the Gurobi solver with default parameters to solve this approximation.

- We approximate Problem (3.37) with the MILP approximation (3.35) in Section 3.3 (i.e., both constraints (3.3b) and (3.3c) are approximated) and use the Gurobi solver with default parameters to solve this approximation.

We report the average performance in Table 3.2. Here, we provide a description of the table headers used in the following tables. The column "N" represents the dimension of the decision variable $\mathbf{x}$, i.e., the number of variables. The column "N_1" represents the number of negative eigenvalues in matrix $\mathbf{Q}_i^n$ of non-convex constraints for all $i \in [m_2]$. The column "Time" indicates the computational time in seconds required to solve each instance. We use "Gap" to represent the relative gap in percentage between a lower bound and an upper bound, as defined in the following equation:

$$\text{Gap} = \frac{\text{upper bound} - \text{lower bound}}{|\text{upper bound}|} \times 100\%. \tag{3.38}$$

To focus on the quality of the lower bound, we unify the upper bounds used in the columns "Gurobi," "LP relaxation," "SOCP relaxation," "SDP relaxation," "A1," and "A2." Specifically, the upper bound used in (3.38) is given by the best-known feasible solution. Note that besides the feasible solution provided by Gurobi, we can also use the "trust-constr" local method in SciPy (Virtanen et al. 2020) to retrieve a feasible solution providing an upper bound. Local methods are commonly used in QCP algorithms (Dong and Luo 2018, Kannan et al. 2022, Bestuzheva et al. 2021) to provide upper bounds. Note also that there may be two values separated by "/" in each of the columns "Gap" and "Time" within the column "Gurobi." It is because we would like to obtain the computational time of Gurobi when it obtains the solution with an optimality gap of 10%. Specifically, we set the Gurobi parameter "MIPGap" to 10%. This setting instructs Gurobi to stop the optimization process when the optimality gap reaches 10%.

Table 3.2: Average Performance of Experiments

		Gurobi		LP Relaxation		SOCP Relaxation		SDP Relaxation		A1		A2	
N	N_1	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
100	1	0%/10.00%	1653/1210	1538.82%	1	571.36%	4	213.08%	113	8.07%	987	3.25%	626
	3	0%/10.00%	1793/1337	1936.25%	1	661.47%	2	247.11%	141	8.34%	1100	3.19%	658
150	1	0%/10.00%	3685/2419	1141.96%	1	354.45%	5	247.82%	496	7.83%	2053	3.35%	1701
	3	0%/10.00%	4745/3081	1529.38%	1	645.70%	2	271.01%	439	9.50%	2484	1.22%	2088
200	1	273.28%	7200	1000.71%	1	434.76%	6	227.32%	1202	9.21%	3752	3.91%	4911
	3	182.91%	7200	1529.38%	1	803.18%	4	209.74%	1387	8.78%	5291	4.26%	5986

3.4.3 Numerical Performance

Note that only relaxing the $\|\mathbf{y}\| \leq t$ part or the $\|\mathbf{z}\| \geq t$ part of Problem (3.3) results in a difficult problem that cannot be solved to optimality within the time limit (i.e., two hours). Therefore, we do not report related results in Table 3.2. Together with the performance of A1 and A2 in Table 3.2, we conclude that the speed-up of approximation comes from the final MILP approximations, and merely approximating the $\|\mathbf{y}\| \leq t$ part or the $\|\mathbf{z}\| \geq t$ part leads to poor performance.

Based on the information provided in Table 3.2, we further have the following observations. First, when $N = 100$ and 150, where the Gurobi solver can solve each instance to optimality within the time limit, our proposed approximations exhibit efficient performance in obtaining lower bounds. On the one hand, our proposed two approximations require less time to provide better lower bounds (with smaller optimality gaps) than the Gurobi solver. For example, when $N = 150$ and $N_1 = 3$, the Gurobi solver takes 3081 seconds to obtain a solution with a gap of 10.00%, but A2 only takes 2088 seconds to obtain a solution with a gap of 1.22%. On the other hand, the two approximations show significant improvement in the optimality gap compared to the LP, SOCP, and SDP relaxations of the original non-convex QCP, where these relaxations require less time to be solved. Moreover, A2 achieves smaller optimality gaps (mostly less than 4%) within a shorter time compared to A1.

Second, when $N = 200$, where the Gurobi solver cannot solve all the instances within the time limit of two hours, our proposed approximations still deliver tight lower bounds.

In such cases, our approximations ensure that the optimality gap remains below 10.00%. For instance, when $N = 200$ and $N_1 = 1$, the Gurobi solver obtains a solution with a gap of 273.28% within two hours, but A1 can obtain a solution with a gap of 9.21% in 3752 seconds. Moreover, when $N = 200$, A2 requires more computational time to be solved compared to A1.

Third, A2 generally outperforms A1 in terms of the optimality gap. However, as N increases, the computational time of A2 increases faster than that of A1. To summarize, our introduced MILP approximations yield better lower bounds within shorter computational times compared to the Gurobi solver when N_1 is small. Note that we do not consider larger N_1 because our approximations may introduce too many binary variables and they are time-consuming to deal with by the solver. The further advancement of the MILP solver can handle this problem and many QCP applications have small N_1 (see Section 3.4.5.1), which justifies the significant applicability of our proposed MILP approximations.

3.4.4 Sensitivity Analyses

To investigate the sensitivity of the approximations to different parameters, we conduct sensitivity analyses by varying the values of m_1, m_2, and ε. In these analyses, we keep N and N_1 fixed at certain values to focus on the impact of the specific parameters being studied. Specifically, we set $N = 100$ and $N_1 = 3$. These analyses provide valuable information on how the performance of the proposed approximations is affected by variations in the parameters m_1, m_2, and ε. For each set of parameters, we generate ten instances and analyze the average performance.

3.4.4.1 Number of Convex Quadratic Constraints m_1

First, we examine the effect of the number of convex quadratic constraints. We consider different values of $m_1 \in \{2, 3, 4, 5, 6\}$ while keeping $m_2 = 2$ and $\varepsilon = 0.01$. We summarize the average performance in Table 3.3.

Our observations indicate that the computational times of both approximations exhibit a slow increase as m_1 increases. This behavior can be attributed to the fact that convex quadratic constraints only introduce linear constraints in the two approximations. In addition, we find that the optimality gap has little correlation with m_1. This suggests that the presence of convex quadratic constraints does not significantly affect the approximation quality.

Table 3.3: Sensitivity Analysis on m_1

	Gurobi		A1		A2	
m_1	Gap	Time	Gap	Time	Gap	Time
2	10.00%	1294	8.93%	842	3.00%	721
3	10.00%	1297	7.69%	934	3.00%	767
4	10.00%	1393	8.34%	1100	3.19%	658
5	10.00%	1313	7.41%	1070	2.78%	801
6	10.00%	1448	8.06%	1139	3.77%	945
average	10.00%	1349	8.09%	1017	3.15%	778

3.4.4.2 Number of Non-convex Quadratic Constraints m_2

Next, we investigate the influence of the number of non-convex quadratic constraints. Similar to the analysis for m_1, we vary m_2 while keeping $m_1 = 4$ and $\varepsilon = 0.01$. We consider $m_2 \in \{1, 2, 3, 4, 5\}$. We summarize the average performance in Table 3.4.

Our observations indicate that the computational times of both approximations increase significantly as m_2 increases. This can be attributed to the fact that non-convex quadratic constraints introduce additional binary variables in our MILP approximations, which pose computational challenges for two approximations. Specifically, the computational time of A2 increases faster than A1. This can be attributed to the fact that A2 generally introduces more complementarity constraints than A1 (see Table 3.1). Similar to our findings with m_1, the optimality gap exhibits little correlation with m_2. This suggests that the presence of non-convex quadratic constraints does not significantly affect the solution quality. Moreover, compared to the Gurobi solver, we know that our approximations are more suitable for a QCP with a few non-convex constraints, which is common in the industry for many QCP problems (see Section 3.4.5).

Table 3.4: Sensitivity Analysis on m_2

	Gurobi		A1		A2	
m_2	Gap	Time	Gap	Time	Gap	Time
1	10.00%	1428	8.60%	648	3.28%	465
2	10.00%	1593	8.34%	1100	3.19%	658
3	10.00%	1885	7.44%	2126	3.43%	2314
4	10.00%	1578	8.63%	3186	3.17%	4260
5	10.00%	1720	8.51%	4308	2.90%	6087
average	10.00%	1641	8.30%	2274	3.19%	2757

3.4.4.3 Quality Parameter ε

Lastly, we examine the impact of the quality parameter ε on the results of our approximations. We test different values of ε while fixing $m_1 = 4$ and $m_2 = 2$. We consider $\varepsilon \in \{0.014, 0.012, 0.01, 0.008, 0.006, 0.004, 0.002\}$ and summarize the average performance in Table 3.5. Note that the performance of the Gurobi solver is not affected by ε. The Gurobi solver needs 1337 seconds to obtain a solution with a gap of 10%.

Our observations reveal that the gap provided by A2 is smaller than that provided by A1. This suggests that A2 yields tighter lower bounds. Meanwhile, for both approximations, the computational time grows rapidly as the value of ε decreases. This can be intuitively explained by the fact that when ε decreases, the number of constraints required to approximate the SOC constraints increases. Consequently, the number of binary variables in each MILP approximation also increases. Therefore, we suggest using a slightly large ε for our approximations because such a ε has little effect on the optimality gap but a significant effect on the computational time.

Table 3.5: Sensitivity Analysis on ε

ε	A1 Gap	A1 Time	A2 Gap	A2 Time
0.014	8.87%	716	3.55%	508
0.012	8.68%	857	3.45%	535
0.010	8.34%	1100	3.19%	658
0.008	7.62%	1886	2.87%	1585
0.006	7.01%	2725	2.56%	2698
0.004	6.38%	3492	2.23%	3935
0.002	5.59%	4625	1.99%	4749
average	7.50%	2200	2.83%	2096

3.4.5 Two Applications

3.4.5.1 Joint Decision and Estimation Problem

As an application of our proposed approximations, we consider the JDE problem, which has been widely applied in the fields of signal processing, wireless communication, power systems, biological information, and medical imaging. In the context of signal processing, Zang and Zhu (2022b) show that the JDE problem can be formulated as:

$$
\begin{aligned}
\min_{j\in\{0,1\}} \ \min_{\mathbf{y}_j,\tau_j} \quad & \mathbf{y}_j^\top \mathbf{c}_j \mathbf{c}_j^\top \mathbf{y}_j && (3.39a)\\
\text{s.t.} \quad & \begin{bmatrix} \mathbf{y}_j & \tau_j \end{bmatrix} \begin{bmatrix} -(\mathbf{A}_j + \mathbf{A}_j^\top) & \mathbf{A}_j \mathbf{e} \\ \mathbf{e}^\top \mathbf{A}_j^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}_j \\ \tau_j \end{bmatrix} \le 0, && (3.39b)\\
& (1-\kappa_j)\tau_j \le 1, && (3.39c)\\
& \mathbf{b}_{1-j}^\top \mathbf{y}_j - \kappa_j \tau_j \le 0, && (3.39d)\\
& \mathbf{b}_j^\top \mathbf{y}_j = 1, && (3.39e)\\
& \mathbf{0} \le \mathbf{y}_j \le \tau_j \mathbf{e}, && (3.39f)\\
& \tau_j \ge 0, && (3.39g)
\end{aligned}
$$

where for each $j \in \{0,1\}$, $\mathbf{b}_j \ge 0$, $\mathbf{c}_j \ge 0$, and $\mathbf{A}_j = \mathbf{b}_j \mathbf{c}_{1-j}^\top - \mathbf{c}_j \mathbf{b}_{1-j}^\top$ are coefficient vectors and matrix and $0 \le \kappa_j \le 1$ is a scalar. Note that the number of negative eigenvalues in the matrix of (3.39b) is usually small (Zang and Zhu 2022a). We consider the epigraph form of Problem (3.39) such that it has the same form as Problem (3.1). In the experiment, we set $\kappa_0 = \kappa_1 = 0.9$ and every element in $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{c}_0$, and $\mathbf{c}_1$ is uniformly generated in the interval $[0, 10]$. We then consider the JDE problems with N decision variables for each $j \in \{0, 1\}$, i.e., variables $\mathbf{y}_j$ and τ_j, where $N \in \{100, 200, 300, 400, 500\}$. For each case, we generate ten instances and summarize the average performance in Table 3.6. Note that ε is set to be 0.01.

Table 3.6: Average Performance of the JDE Problem

N	Gurobi Gap	Gurobi Time	A1 Gap	A1 Time	A2 Gap	A2 Time
100	10.00%	1018	7.34%	279	3.07%	163
200	10.00%	1832	7.30%	308	3.64%	142
300	10.00%	4745	7.39%	515	3.39%	222
400	43.19%	7200	8.08%	603	4.11%	443
500	82.82%	7200	7.29%	710	3.68%	623
average	31.20%	4399	7.48%	483	3.58%	319

3.4.5.2 Two-trust-region Subproblem

The TTRS is a non-convex quadratic problem with two ellipsoidal constraints, i.e.,

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{c}^\top \mathbf{x} && (3.40a)\\
\text{s.t.} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \le d_1, && (3.40b)\\
& \|\mathbf{x}\| \le d_2, && (3.40c)
\end{aligned}
$$

where $\mathbf{c}$, $\mathbf{b}$, $\mathbf{Q}$, and $\mathbf{A}$ are coefficient vectors and matrices and d_1 and d_2 are two scalars. Similar to Problem (3.39), we consider the epigraph form of Problem (3.40) such that it has the same form as Problem (3.1). Note that Problem (3.40) has two convex quadratic constraints and one non-convex quadratic constraint. In this experiment, we follow Section 5.3 of Burer and Anstreicher (2013) to randomly generate instances. The number of negative eigenvalues in $\mathbf{Q}$ is randomly chosen from $\{1, 2, 3\}$. We then consider the TTRS problems with N decision variables, i.e., $\mathbf{x} \in \mathbb{R}^N$, where $N \in \{100, 200, 300, 400, 500\}$. For each case, we generate ten instances and summarize the average performance in Table 3.7. Note that ε is set to 0.01.

Table 3.7: Average Performance of the TTRS Problem

	Gurobi		A1		A2	
N	Gap	Time	Gap	Time	Gap	Time
100	10.00%	2295	7.31%	519	3.01%	369
200	10.00%	2820	6.75%	608	3.53%	429
300	10.00%	4124	7.54%	666	3.02%	489
400	10.00%	5510	6.88%	787	2.94%	669
500	49.27%	7200	7.14%	971	3.23%	781
average	17.85%	4390	7.12%	710	3.15%	547

3.4.5.3 Numerical Performance

Our observations indicate that the computational time of the Gurobi solver increases significantly as the value of N increases. In contrast, the computational times of our proposed two approximations increase at a slower rate with increasing N. Specifically, even for relatively large values of N, such as $N = 500$, the two approximations can obtain solutions with much smaller gaps within 1000 seconds, whereas the Gurobi solver requires two hours to obtain solutions with a larger gap. This highlights the advantage of our approximations in terms of optimality gaps. Furthermore, our findings demonstrate that A2 outperforms A1 in the context of the JDE problem and the TTRS problem. The approximation A2 consistently yields smaller values for the optimality gap and computational time than A1.

3.5 Conclusion

A non-convex QCP optimizes a linear objective function subject to non-convex quadratic constraints. Despite its widespread applications in various industries and academia, finding an optimal solution for a non-convex QCP remains a challenging task. Currently, most QCP solvers utilize the spatial branch-and-bound algorithm to search for an optimal solution. One of the most important components of this algorithm is deriving efficient lower bounds, which help reduce the size of the branch-and-bound tree. In this chapter, we propose two MILP approximations to provide tighter lower bounds for a non-convex QCP. To the best of our knowledge, we are among the first to develop asymptotically tight MILP approximations for a non-convex QCP.

Initially, considering a non-convex QCP, we employ an eigenvalue-based decomposition to reformulate non-convex quadratic constraints. Specifically, we transform each non-convex quadratic constraint into an SOC constraint $\|\mathbf{y}\| \leq t$ and the complement of an SOC constraint $\|\mathbf{z}\| \geq t$. Subsequently, we introduce a polyhedral approximation for $\|\mathbf{y}\| \leq t$. Based on the approximation of $\|\mathbf{y}\| \leq t$, we develop an approximation for $\|\mathbf{z}\| \geq t$. Combining these approximation results, we obtain an LPCC approximation for the non-convex QCP. Furthermore, we establish that the feasible solution of the LPCC satisfies the "ε-relaxed" constraints of the non-convex QCP (Proposition 9). More importantly, we demonstrate that the optimal value of the LPCC converges to the optimal value of the non-convex QCP asymptotically (Theorem 7). Finally, we show the boundedness of the LPCC when the original non-convex QCP is bounded (Section 3.2.5). As a result, we reformulate the LPCC into an MILP by utilizing this boundedness property. Motivated by the trade-off between the number of variables and the number of constraints in the context of extended formulations for approximations, we provide an alternative MILP approximation in Section 3.3, where the asymptotic convergence property also holds. Compared to the first approximation, the second approximation typically introduces fewer variables but may result in more constraints when the approximation quality parameter ϵ is small.

We conduct extensive computational experiments to validate the effectiveness of our proposed approximations over our randomly generated instances and the unitbox instances from the literature. The numerical results demonstrate that the proposed approximations yield smaller optimality gaps within shorter computational times than a state-of-the-art commercial solver. Meanwhile, our proposed approximations show a significant improvement in the optimality gap compared to the LP, SOCP, and SDP relaxations of the original non-convex QCP. We also apply our approximations to two problems: the joint decision and estimation problem and the two-trust-region subproblem. The numerical results highlight the significant advantages of our proposed approximations in terms of computational time and optimality gap.

Chapter 4

Data-Driven Chance-Constrained Planning for Distributed Generation: A Partial Sampling Approach

4.1 Introduction

With technological development and governments' support, renewable energy has drawn significant attention and investment. A widely used strategy to exploit renewable energy is to integrate renewable distributed generation (RDG) units into existing power distribution grids. Correct installation of RDG units can help power system operators provide customers with affordable and reliable energy, while improper placement may result in many problems, e.g., system instability and power losses (Keane et al. 2013), due to the intermittency of renewable energy. Such intermittency from RDG units may lead to cascading problems such as an imbalance of electricity supply and demand and system blackouts (Georgilakis and Hatziargyriou 2013). Therefore, proper siting (i.e., location) and sizing (i.e., capacity) decisions of the RDG units are of great significance for power system operators to ensure the benefits of renewable energy and maintain reliable operations of the power distribution grids.

Besides RDG units, energy storage (ES) has also been considered for use in power distribution grids. It is because ES units can provide a buffer against an imbalance of supply and demand, thereby reducing operating costs and increasing a power distribution grid's probability of meeting demand. Such benefits may offset the installation and operating costs of ES units and even lead to profits. For example, Abbey and Joos (2007) adopts ES to help support wind energy applications and Barton and Infield (2004) uses ES to increase the penetration of more general renewable generation by smoothing out the effects of intermittency. The positive results from these studies indicate the necessity to integrate ES units into a power distribution grid with RDG units. Thus, it is important to determine the optimal siting and sizing decisions of both the RDG and ES units.

For ease of exposition, we refer to the problem of siting and sizing both RDG and ES units in a power distribution grid as *the planning problem* thereafter. Such a problem largely relies on accurate power flow analysis. Two mathematical models are often considered to calculate the optimal power flow in power grids: the direct current optimal power flow (DCOPF) and the alternating current optimal power flow (ACOPF) models. The DCOPF model is composed of linear constraints and thus is easier to solve, while it oversimplifies the physical features. In contrast, the ACOPF model is relatively accurate by considering active and reactive power-generation limits, demand limits, bus voltage limits, and network flow limits (Kuang et al. 2016), while it is nonlinear and nonconvex. The DCOPF model provides a linear approximation of the ACOPF model, with systematic errors though. Such inaccuracy is acceptable for large-scale power networks but is often unacceptable for local distribution grids. Therefore, this chapter uses the ACOPF model to accurately simulate a power distribution grid.

Different variants of the planning problem have recently drawn much attention from academia and industry. Ehsan and Yang (2018) provides a review of related studies on the RDG planning in the power distribution network, and Oree et al. (2017) reviews related studies on more general generation expansion planning. Specifically, Celli et al. (2005) plans the locations and capacity sizes of RDG units based on simplified load flow calculation in a multiobjective optimization model that is further solved by a genetic algorithm, while ES units are not considered. Similarly, without ES units included, Melgar-Dominguez et al. (2018) considers the siting and sizing of RDG units in a power distribution grid in a two-stage robust optimization model. Pandžić et al. (2014) makes the siting and sizing decisions for the ES units in a power transmission grid via a three-stage mixed-integer linear program, while the DCOPF model is adopted. A recent work can be found in Fathabad et al. (2020), which considers only the sizing of RDG units in a two-stage distributionally robust optimization model and makes the siting decisions via sensitivity analyses. The above existing efforts demonstrate the significance of siting and sizing RDG and ES units.

However, the large-scale installation of RDG units adds significant uncertainties to a power distribution grid due to the intermittency of renewable energy, requiring methodological innovation to deal with the already challenging operations of a power distribution grid. To that end, many stochastic programming models dealing with uncertainties have been developed to support effective distribution grid operations. For example, Zhou et al. (2013) proposes a two-stage stochastic programming model for the optimal planning of distributed energy systems under demand and supply uncertainties. Tan et al. (2014) proposes a two-stage stochastic programming-based optimal power flow model for the operation of distribution networks with uncertainties from wind power.

In this chapter, we adopt chance-constrained programming to model the planning and operational decisions under uncertainties, including renewable generation and load uncertainties, to ensure the feasibility of the distribution system under a high probability. Specifically, we note that, due to such uncertainties, a power distribution grid may face various reliability issues. For instance, power outages often happen when the power supply is insufficient, such as facing a natural disaster like a typhoon. A load bus that is far away from the upper stream grid may not receive the power injected from the upper stream grid or local distributed generators because of distribution line capacity limit and line loss. The distribution line contingency may also happen. Therefore, to help relieve such pressure and ensure system reliability, we build a two-stage chance-constrained (TCC) model (where the siting and sizing decisions are in the first stage and operational decisions are in the second stage) to solve *the planning problem*. More importantly, considering various reliability issues, we adopt a chance constraint to ensure *all* the operational constraints in the second stage are satisfied simultaneously, rather than ensuring load satisfaction only. As such, our proposed TCC model becomes extremely difficult to solve, requiring further innovation in solution approaches.

We note that the chance-constrained model is a risk-averse decision-making tool that can help grid operators actively control the probability of unfavorable outcomes (e.g., system blackouts). The chance-constrained model is widely used for power system operations. For instance, Roald and Andersson (2017) solves the chance-constrained ACOPF problems to ensure that operational constraints are satisfied with the desired probability. Lubin et al. (2019) further provides convex approximations via second-order conic programming, and Fathabad et al. (2023) provides asymptotically tight conic approximations for the chance-constrained ACOPF problems. Similarly, Ozturk et al. (2004) studies the chance-constrained unit commitment formulations, and Wu et al. (2014) investigates the chance-constrained day-ahead scheduling. Several studies use TCC models, which are much more complex than single-stage chance-constrained models. For instance, Wang et al. (2012) is among the first to propose a TCC model for the unit commitment problem, while it considers a single chance constraint to ensure load satisfaction only. Zhang et al. (2017b) considers a similar TCC model for the unit commitment problem while presenting a bilinear mixed-integer reformulation solved by Benders decomposition following the study in Zeng et al. (2014). Pozo and Contreras (2012) formulates the chance constraints based on the definition of conditional value at risk in a TCC model for the unit commitment problem and reformulates these constraints using sampling-based approaches.

Wu et al. (2016) specifically considers wind uncertainty in the chance-constrained unit commitment model. However, to the best of our knowledge, few studies apply the TCC models in the planning of RDG and ES units together in a distribution system considering the ACOPF model. More importantly, different from the existing studies above, this chapter considers all the operational constraints simultaneously in a difficult joint chance constraint.

Nevertheless, the TCC model is intractable in general, specifically when the random parameters follow an underlying continuous (yet unknown) probability distribution and are of high dimensions, posing severe computational challenges (Nemirovski and Shapiro 2007). The existing studies primarily adopt two approximation approaches to solve a chance-constrained model: a convex (or tractable) approximation approach (Lubin et al. 2019) and a sampling-based approach (Wang et al. 2012). The former is not applicable to a TCC model because the second-stage recourse decision is a function of the first-stage decision and random parameters and it is impossible to algebraically characterize this function. Thus, this chapter uses the latter. Specifically, we propose two sampling techniques to reformulate our TCC model: the standard sample average approximation (SAA) method and a new partial sample average approximation (PSAA) method. Both approximations lead to mixed-integer convex quadratic programs. However, the SAA introduces many additional binary variables corresponding to the samples, creating computational complexity to the already challenging distributed energy resource planning problem. In contrast, the PSAA samples only a part of the random parameters, and we use a non-parametric estimation method to approximate the probability distribution of the remainder, leading to an efficient data-driven approach. We only need to introduce additional continuous variables corresponding to the samples to reformulate the model, reducing the computational complexity. Thus, the reformulation can be scaled up to solve large-scale instances.

The main contributions of the chapter are as follows.

- We develop a novel TCC model for the distributed energy resource planning problem that considers both the placement and capacity of RDG and ES units under uncertainty, combined with the ACOPF model, in a distribution grid with multiple periods. We consider all the operational constraints in the second stage to be satisfied simultaneously in a joint chance constraint.

- We are the first to develop the PSAA approach using historical data to solve the above TCC model for a significant industry problem. We extend the PSAA idea that solves single-stage chance-constrained models.

- Our extensive experiments on the IEEE 33-Bus and 123-Bus systems using real data show that the PSAA approach performs better than the standard SAA approach because the former provides better solutions in a shorter time in in-sample tests and provides better guaranteed probability for system reliability in out-of-sample tests. The effectiveness of ES units in reducing total costs and improving system balance is also demonstrated.

The remainder of this chapter is organized as follows. A two-stage model and its TCC counterpart for the planning problem are presented in Section 4.2. The TCC model is then approximated using the SAA and PSAA methods in Section 4.3. We provide computational results and explanations in Section 4.4. Section 4.5 concludes this chapter. The notation used throughout this chapter is summarized in Appendix C.1.

4.2 Mathematical Model

In this section, we present a two-stage stochastic programming model and its TCC counterpart for the planning problem.

4.2.1 A Two-Stage Model

We focus on a typical distribution grid topology: the radial network. Such a network has a tree structure and connects to a transmission network via a single bus (Bus 0). In our distribution network, the power supply comes from four sources: the transmission network, the traditional dispatchable distributed generation (DDG) units and reactive sources, the RDG units, and the ES units. The last three sources are located in some buses of the distribution grid. While the DDG units and reactive sources have already been placed (i.e., given system input data), the RDG and ES units are to be installed (i.e., system decision variables). The operating costs for the supply sources include the payment to the transmission network, the cost of power generation from the DDG units, and the cost of charging/discharging the ES units. Here we investigate the optimal siting and sizing of $\bar{K}$ candidate RDG units and R candidate ES units in a distribution grid with buses $\mathcal{N}$, to minimize the total cost across the planning horizon. The cost includes deterministic investment/maintenance costs and stochastic operating costs (due to the uncertainties in load and renewable power generation). In the following, we formulate the planning problem as a two-stage optimization model and describe the corresponding first-stage and second-stage objectives and constraints.

1) **First-stage model**: The first-stage objective minimizes the total cost of building, maintaining, and operating the RDG and ES units, with the model formulated as follows.

$$\min_{\mathbf{\Omega}^1} \; C_1\left(\mathbf{\Omega}^1\right) + \mathbb{E}\left[Q\left(\mathbf{\Omega}^1, \boldsymbol{\xi}\right)\right] \tag{4.1a}$$

$$\text{s.t.} \; \sum_{n\in\mathcal{N}} z_{kn} \leq 1, \; \forall\, k \in [K], \tag{4.1b}$$

$$\sum_{k=1}^{K}\sum_{n\in\mathcal{N}} z_{kn} \leq \bar{K}, \tag{4.1c}$$

$$x_k = \sum_{l=1}^{L} u_{kl}\bar{x}_l, \; \forall\, k \in [K], \tag{4.1d}$$

$$\sum_{l=1}^{L} u_{kl} = \sum_{n\in\mathcal{N}} z_{kn}, \; \forall\, k \in [K], \tag{4.1e}$$

$$\sum_{n\in\mathcal{N}} w_{rn} \le 1,\ \forall\, r \in [R], \tag{4.1f}$$

$$\sum_{n\in\mathcal{N}} w_{rn}\underline{y}_r \le y_r \le \sum_{n\in\mathcal{N}} w_{rn}\bar{y}_r,\ \forall\, r \in [R], \tag{4.1g}$$

where $\boldsymbol{\Omega}^1 := [\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{y}]^\top$ is the vector of first-stage variables. The first part in the objective function (4.1a)

$$C_1\left(\boldsymbol{\Omega}^1\right) := \sum_{k=1}^{K}\left(\sum_{n\in\mathcal{N}} c^0_{kn} z_{kn} + \left(c^1_k + Tc^2_k\right) x_k\right) + \sum_{r=1}^{R}\left(\sum_{n\in\mathcal{N}} d^0_{rn} w_{rn} + \left(d^1_r + Td^2_r\right) y_r\right),$$

represents the total deterministic cost, including the setup costs and the size-based investment/maintenance costs of the RDG and ES units. Specifically, $C_1\left(\boldsymbol{\Omega}^1\right)$ includes two parts: (i) the setup costs of the RDG and ES units, $\sum_{k=1}^{K}\left(\sum_{n\in\mathcal{N}} c^0_{kn} z_{kn}\right) + \sum_{r=1}^{R}(\sum_{n\in\mathcal{N}} d^0_{rn}$ $w_{rn})$, and (ii) the investment/maintenance costs of the RDG and ES units, $\sum_{k=1}^{K}((c^1_k + Tc^2_k$ $)x_k) + \sum_{r=1}^{R}\left(\left(d^1_r + Td^2_r\right) y_r\right)$. In part (i), z_{kn} is a binary variable indicating whether the kth candidate RDG unit is located at bus n and w_{rn} is a binary variable indicating whether the rth candidate ES unit is located at Bus n. In part (ii), x_k and y_r are continuous variables indicating the capacities of kth candidate RDG unit and rth candidate ES unit, respectively.

The second part in the objective function (4.1a), $\mathbb{E}[Q(\boldsymbol{\Omega}^1, \boldsymbol{\xi})]$, represents the expected minimum operating costs over all T periods, which is defined explicitly in (4.2). Note that $\boldsymbol{\xi}$ is defined as the vector of all random variables (see Table C.1 for the details), with $(\Xi, \mathcal{F}, \mathbb{P})$ denoting its probability space. Regarding the constraints (4.1b) - (4.1g), they link the variables z_{kn}, w_{rn}, x_k, y_r, and u_{kl}, by which the setup costs are linked with the investment/maintenance cost. Specifically, constraints (4.1b) show that a given kth candidate RDG unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraint (4.1c) enforces that the total number of RDG units installed should not exceed the limit $\bar{K}$.

In constraints (4.1d) and (4.1e), u_{kl} is a binary variable indicating whether the capacity of the kth candidate RDG unit is the lth element in $\mathcal{X}$, $\bar{x}_l$. Thus, constraints (4.1d) and

(4.1e) ensure that if a given kth candidate RDG unit is installed, then the capacity of this RDG unit should be one of the pre-defined values in $\mathcal{X}$. For any given kth candidate RDG unit, the summation $\sum_{l=1}^{L} u_{kl}$ is equivalent to the $\sum_{n\in\mathcal{N}} z_{kn}$, which may take only the value 0 or 1. Such equivalence is not related to the locations of candidate RDG units. Constraints (4.1d) and (4.1e) are motivated by the practice where various regions and institutions have different regulations on the capacity of distributed generation, and thus the capacity of RDG units vary (Ackermann et al. 2001).

Similar to (4.1b), constraints (4.1f) show that a given rth candidate ES unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraints (4.1g) show that if a given rth candidate ES unit is installed (i.e., $w_{rn} = 1$ for a bus n), then its capacity should be between the lower bound $\underline{y}_r$ and the upper bound $\bar{y}_r$. Note that multiple new assets (including RDG and ES units) may be installed eventually.

From the above constraints, we can observe that z_{kn} impacts x_k and w_{rn} impacts y_r, i.e., the binary variables indicating the location of new assets have an impact on the investment/maintenance cost.

2) **Second-stage model**: Given a first-stage decision $\mathbf{\Omega}^1$ and a realization $\boldsymbol{\xi}$ of the uncertain load and renewable generation, the second-stage objective minimizes the distribution grid's operating costs $Q(\mathbf{\Omega}^1, \boldsymbol{\xi})$, where $\boldsymbol{\xi} := [\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^T]^\top$, while respecting a set of physical constraints such as the ACOPF constraints. The operating costs include the cost of purchasing active/reactive energy via Bus 0, the cost of fuel used and emissions created in generating active power in the DDG units, and the cost of charging and discharging the stored energy. The operating costs also include the load-shedding variables LS^t_{1mn}, LS^t_{2mn}, and a penalty factor p to account for any unsatisfied load. Meanwhile, note that introducing the load-shedding makes the resulting two-stage stochastic programming model satisfy the complete recourse assumption. Let $\mathbf{\Omega}^2$ be the vector of all second-stage

variables and let

$$C_2\left(\mathbf{\Omega}^2\right) := \sum_{t=1}^{T}\left(c_p^t p_0^t + c_q^t q_0^t + \sum_{n\in\mathcal{B}_1} c_n^f p_n^t + \sum_{n\in\mathcal{B}_1} c_n^e \omega p_n^t \right.$$
$$\left. + \sum_{r=1}^{R} e_1 f_r^t + \sum_{r=1}^{R} e_2 g_r^t + \sum_{(m,n)\in\mathcal{E}} p\left(LS_{1mn}^t + LS_{2mn}^t\right)\right)$$

be the total operating costs in the second stage. The second-stage problem, whose optimal value is denoted by $Q(\mathbf{\Omega}^1, \boldsymbol{\xi})$, can be formulated as follows. (For ease of exposition, all constraints with a superscript t hold for all $t \in [T]$.)

$$\min_{\mathbf{\Omega}^2} \quad C_2\left(\mathbf{\Omega}^2\right) \tag{4.2a}$$

$$\text{s.t.} \quad \underline{p}_n^t \le p_n^t \le \bar{p}_n^t,\ \forall\, n \in \mathcal{B}_1, \tag{4.2b}$$

$$\underline{q}_n^t \le q_n^t \le \bar{q}_n^t,\ \forall\, n \in \mathcal{B}_2, \tag{4.2c}$$

$$\underline{v} \le \left|V_n^t\right|^2 \le \bar{v},\ \forall\, n \in \mathcal{N}\setminus\{0\}, \tag{4.2d}$$

$$p_0^t = \sum_{n\in\mathcal{B}_0} P_{0n}^t, \quad q_0^t = \sum_{n\in\mathcal{B}_0} Q_{0n}^t, \tag{4.2e}$$

$$P_{mn}^t - \mathfrak{R}_{mn}\left|I_{mn}^t\right|^2 + LS_{1mn}^t = d_{pn}^t - \sum_{k=1}^{K} z_{kn} s_k^t x_k - \delta_n p_n^t + \sum_{l\in\mathcal{N}_n} P_{nl}^t + \sum_{r=1}^{R} w_{rn}\left(f_r^t - g_r^t\right),$$
$$\forall\,(m,n)\in\mathcal{E}, \tag{4.2f}$$

$$Q_{mn}^t - \mathfrak{X}_{mn}\left|I_{mn}^t\right|^2 + LS_{2mn}^t = d_{qn}^t - \tau_n q_n^t + \sum_{l\in\mathcal{N}_n} Q_{nl}^t,\ \forall\,(m,n)\in\mathcal{E}, \tag{4.2g}$$

$$b_r^0 = 0,\ \forall\, r \in [R], \tag{4.2h}$$

$$0 \le b_r^t \le y_r,\ \forall\, r\in[R], \tag{4.2i}$$

$$b_r^t - b_r^{t-1} = \gamma f_r^t - g_r^t/\gamma,\ \forall\, r\in[R], \tag{4.2j}$$

$$\left|V_m^t\right|^2 - \left|V_n^t\right|^2 = 2\mathfrak{R}_{mn}P_{mn}^t + 2\mathfrak{X}_{mn}Q_{mn}^t - \left(\mathfrak{R}_{mn}^2 + \mathfrak{X}_{mn}^2\right)\left|I_{mn}^t\right|^2,\ \forall\,(m,n)\in\mathcal{E}, \tag{4.2k}$$

$$\left\|\left[2P_{mn}^t, 2Q_{mn}^t, \left|V_m^t\right|^2 - \left|I_{mn}^t\right|^2\right]\right\|_2 \le \left|V_m^t\right|^2 + \left|I_{mn}^t\right|^2, \forall\,(m,n)\in\mathcal{E}, \tag{4.2l}$$

$$\left(P_{mn}^t\right)^2 + \left(Q_{mn}^t\right)^2 \le (LC_{mn})^2,\ \forall\,(m,n)\in\mathcal{E}, \tag{4.2m}$$

$$p_0^t \ge 0,\ q_0^t \ge 0, \tag{4.2n}$$

$$LS_{1mn}^t \ge 0,\ LS_{2mn}^t \ge 0,\ \forall\,(m,n)\in\mathcal{E}. \tag{4.2o}$$

Here, $\mathbf{\Omega}^2$ consists of p_0^t, q_0^t for $t \in [T]$, $p_n^t, q_n^t, V_n^t, P_{0n}^t, Q_{0n}^t$ for $n \in \mathcal{N}$, $t \in [T]$, f_r^t, g_r^t, b_r^t for $r \in [R], t \in [T]$, P_{nl}^t, Q_{nl}^t for n such that $(m,n) \in \mathcal{E}$ for some m, $l \in \mathcal{N}_n$, $t \in [T]$, and $P_{mn}^t, Q_{mn}^t, I_{mn}^t, LS_{1mn}^t, LS_{2mn}^t$ for $(m,n) \in \mathcal{E}, t \in [T]$.

We explain all the constraints in the model (4.2) as follows. The power generated by the DDG units and the reactive sources is bounded by (4.2b) and (4.2c), respectively. Constraint (4.2d) sets the bounds on the voltage of each bus. Constraint (4.2e) represents the active and reactive balance equations at Bus 0. Constraints (4.2f) and (4.2g) are active and reactive power balance equations from Kirchhoff's current law. The following Fig. 4.1 illustrates the active power flow balance for each distribution line $(m,n) \in \mathcal{E}$. The

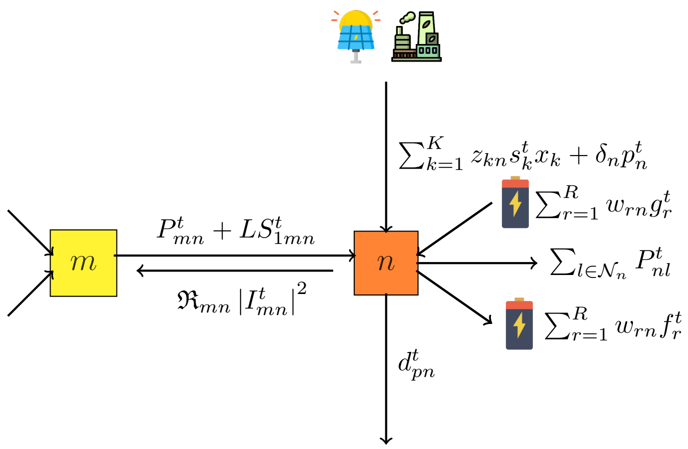

Figure 4.1: Power Flow Blance.

balance of power in storage is initialized by (4.2h) and bounded by (4.2i). ES balance between two consecutive periods is shown in (4.2j), considering ES charging/discharging efficiency. Constraint (4.2k) represents the voltage drop on each line. Constraint (4.2l) is the branch power-flow constraint, and the capacity of each distribution line is limited by (4.2m). Nonnegativity constraints are listed in (4.2n) and (4.2o).

Note that the load-shedding variables LS_{1mn}^t and LS_{2mn}^t are defined over each distribution line $(m,n) \in \mathcal{E}$. Once the model is solved, one can also compute the load-shedding values at each bus easily. We also note that the branch flow model is applied here to formulate the ACOPF constraints in (4.2f) – (4.2g) and (4.2k) – (4.2m), where (4.2l) includes a set of a second-order conic (SOC) constraints. These constraints represent the ACOPF constraints via convex relaxation following the study in Farivar and Low (2013). Specifically, Farivar and Low (2013) removes the voltage and current angles while introducing squared voltage and current magnitudes, and relaxes the nonconvex quadratic constraints with convex SOC constraints. More importantly, Farivar and Low (2013) shows that the obtained convex relaxation is exact when the distribution network is radial. As most practical distribution networks are radial grids (Farivar and Low 2013), we also consider a radial network in this chapter. Therefore, the constraints (4.2f) – (4.2g) and (4.2k) – (4.2m) form an exact reformulation of the ACOPF constraints. As such, we obtain a second-order conic programming (SOCP) formulation in (4.2), which enables large-scale applications due to the computational efficiency of SOCP formulations. We note that there are also other types of approximations to formulate the ACOPF constraints, such as linearized distribution flow (LinDistFlow) (Baran and Wu 1989), where voltage drop and line power flows are approximately linearly related to power injections.

We summarize the two-stage stochastic programming model of the planning problem as

$$\begin{aligned} \min_{\boldsymbol{\Omega}^1, \boldsymbol{\Omega}^2(\cdot)} \quad & C_1\left(\boldsymbol{\Omega}^1\right) + \mathbb{E}\left[C_2\left(\boldsymbol{\Omega}^2\right)\right] \\ \text{s.t.} \quad & (4.1\text{b}) - (4.1\text{g}), (4.2\text{b}) - (4.2\text{o}). \end{aligned} \tag{4.3}$$

Note that $\boldsymbol{\Omega}^2$ and constraints (4.2b) – (4.2o) are dependent on $\boldsymbol{\xi}$. As the first-stage problem (4.1) is an integer program and the second-stage problem (4.2) is an SOCP, the entire two-stage problem (4.3) is a mixed-integer SOCP.

4.2.2 A TCC Model

In the two-stage stochastic programming model (4.3), we allow the load to be unsatisfied with a penalty cost, and the goal is to choose the best first-stage decision Ω^1 such that the total first-stage and second-stage cost can be minimized. However, the power system operators often need to control the reliability of the power system, which is measured as the probability of load satisfaction. Therefore, in this section, we use the chance constraints to ensure this requirement and propose a two-stage chance-constrained model.

First, let us revisit the two-stage stochastic programming model (4.3) here:

$$\begin{aligned}&\min_{\Omega^1} C_1\left(\boldsymbol{\Omega}^1\right)+\mathbb{E}\left[Q\left(\Omega^1,\boldsymbol{\xi}\right)\right]\\&\text{s.t. } (4.1\text{b})-(4.1\text{g}),\end{aligned}$$

where $Q(\Omega^1,\boldsymbol{\xi}) = \min_{\Omega^2}\{C_2(\Omega^2) \mid (4.2\text{b}) - (4.2\text{o})\}$. Note that this two-stage stochastic programming model satisfies the complete recourse assumption because the load-shedding variables LS^t_{1mn} and LS^t_{2mn} in Ω^2 can be seen as slack variables, which are penalized in the objective function $C_2(\Omega^2)$.

Second, for any feasible Ω^1 and the realization of $\boldsymbol{\xi}$, "load satisfaction" means that there exists Ω^2 can satisfy (4.2b) − (4.2o) and $LS^t_{1mn} = LS^t_{2mn} = 0,\ \forall\,(m,n)\in\mathcal{E},\ t\in[T]$. For the ease of presentation, we denote another second-stage problem:

$$\begin{aligned}\bar{Q}(\Omega^1,\boldsymbol{\xi}) = \min_{\Omega^2}\quad & C_2(\Omega^2)\\ \text{s.t.}\quad & (4.2\text{b})-(4.2\text{o}),\\ & LS^t_{1mn} = LS^t_{2mn} = 0,\ \forall\,(m,n)\in\mathcal{E},\ t\in[T]. \qquad (4.4)\end{aligned}$$

Note that $\bar{Q}(\Omega^1,\boldsymbol{\xi})$ has extra constraints (4.4) compared with $Q(\Omega^1,\boldsymbol{\xi})$. Meanwhile, $Q(\Omega^1,\boldsymbol{\xi})$ can be seen as the recovery operation problem (Liu et al. 2016). The main difference between the two problems is that, given any Ω^1 and $\boldsymbol{\xi}$, Problem $Q(\Omega^1,\boldsymbol{\xi})$ can always be satisfied, but Problem $\bar{Q}(\Omega^1,\boldsymbol{\xi})$ may not.

Third, we denote the following new two-stage stochastic programming model:

$$\min_{\Omega^1} C_1\left(\boldsymbol{\Omega}^1\right)+\mathbb{E}\left[\bar{Q}\left(\Omega^1,\boldsymbol{\xi}\right)\right] \qquad (4.5)$$

$$\text{s.t. } (4.1b) - (4.1g).$$

Meanwhile, for any $\boldsymbol{\xi} \in \Xi$, we define:

$$\mathbf{P}(\boldsymbol{\xi}) = \{\Omega^1 \text{ satisfies } (4.1b) - (4.1g) \mid \exists\, \Omega^2 \text{ satisfies } (4.2b) - (4.2o) \text{ and } (4.4)\}.$$

Note that Problem (4.5) implicitly enforces that $\Omega^1 \in \mathbf{P}(\boldsymbol{\xi})$ for almost every $\boldsymbol{\xi} \in \Xi$, since otherwise the objective value is infinite. Although Problem (4.5) can ensure the load satisfaction with probability one, it may lead to an infeasible model or may lead to a model in which the first-stage solutions are too costly.

Thus, in this section, the constraint that the second-stage must be feasible in all possible realizations is replaced with the relaxed version that enforces this to hold with high probability, i.e.,

$$\mathbb{P}\left(\boldsymbol{\Omega}^2 \text{ in } (4.2) \text{ satisfies } (4.4) \,|\, \boldsymbol{\Omega}^1\right) \geq 1 - \eta. \tag{4.6}$$

In other words, $\mathbb{P}(\Omega^1 \in \mathbf{P}(\boldsymbol{\xi})) \geq 1 - \eta$. As a result, we can follow Liu et al. (2016) to provide the following TCC model (two-stage chance-constrained programs with recovery):

$$\begin{aligned}
\min_{\Omega^1, \mathbf{I}} \quad & C_1\left(\Omega^1\right) + \mathbb{P}\left(\mathbf{I}_{\boldsymbol{\xi}} = 0\right) \mathbb{E}\left[\bar{Q}(\Omega^1, \boldsymbol{\xi}) \mid \mathbf{I}_{\boldsymbol{\xi}} = 0\right] + \mathbb{P}\left(\mathbf{I}_{\boldsymbol{\xi}} = 1\right) \mathbb{E}\left[Q(\Omega^1, \boldsymbol{\xi}) \mid \mathbf{I}_{\boldsymbol{\xi}} = 1\right] \qquad (4.7)\\
\text{s.t.} \quad & (4.1b) - (4.1g), (4.6),\\
& \mathbf{I}_{\boldsymbol{\xi}} = 0 \Longrightarrow \Omega^1 \in \mathbf{P}(\boldsymbol{\xi}), \quad \forall \boldsymbol{\xi} \in \Xi,\\
& \mathbf{I}_{\boldsymbol{\xi}} \in \{0, 1\}, \quad \forall \boldsymbol{\xi} \in \Xi,
\end{aligned}$$

where $\mathbf{I}_{\boldsymbol{\xi}}$ are indicator decision variables. However, the above problem (4.7) is difficult to solve. Specifically, constraint (4.6) is nonconvex and thus problem (4.7) becomes intractable. Next, we introduce two approximation methods to address this challenge.

4.3 Solution Approaches

In this section, we describe two approximation methods for solving the problem (4.7): the SAA and PSAA methods.

4.3.1 The SAA Formulation

The SAA is a classic sampling technique that is widely used in chance-constrained problems. It approximates the expectation of random variables using their sample means. The probability of an event E can be reformulated as an expectation as follows:

$$\mathbb{P}(E) = \mathbb{E}\left[\mathbb{I}(E)\right],$$

where $\mathbb{I}(\cdot)$ is an indicator function that takes a value of 1 when the event happens and 0 otherwise. Let Π_1 be the total number of samples of the random vector $\boldsymbol{\xi}$ and $[\Pi_1]$ be the set of all the samples, and let $\boldsymbol{\xi}_\pi$ be a specific sample for any $\pi \in [\Pi_1]$. The SAA approximates $\mathbb{P}\left\{\boldsymbol{\Omega}^2 \text{ in (4.2) satisfies (4.4)} \,|\, \boldsymbol{\Omega}^1\right\}$ in (4.6) with

$$\frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} \mathbb{I}\left(\boldsymbol{\Omega}^2_\pi \text{ satisfies (4.2b)} - \text{(4.2o)}, \text{(4.4)} \,|\, \boldsymbol{\Omega}^1\right),$$

where $\boldsymbol{\Omega}^2_\pi$ is a copy of the second-stage variables $\boldsymbol{\Omega}^2$ corresponding to $\boldsymbol{\xi}_\pi$ for each $\pi \in [\Pi_1]$. Let $\text{(4.2b)}_\pi - \text{(4.2o)}_\pi$ be constraints (4.2b) – (4.2o) with $\boldsymbol{\xi}$ replaced by $\boldsymbol{\xi}_\pi$ and $\boldsymbol{\Omega}^2$ replaced by $\boldsymbol{\Omega}^2_\pi$. The SAA approximation of (4.7) thus becomes:

$$\begin{aligned} \min_{\boldsymbol{\Omega}^1,\, \boldsymbol{\Omega}^2_\pi, \forall \pi \in [\Pi_1]} \quad & C_1\left(\boldsymbol{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\boldsymbol{\Omega}^2_\pi\right) \\ \text{s.t.} \quad & \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} \mathbb{I}\left(\boldsymbol{\Omega}^2_\pi \text{ satisfies (4.2b)} - \text{(4.2o)}, \text{(4.4)} \,|\, \boldsymbol{\Omega}^1\right) \geq 1 - \eta. \\ & \text{(4.1b)} - \text{(4.1g)},\ \text{(4.2b)}_\pi - \text{(4.2o)}_\pi,\ \forall \pi \in [\Pi_1]. \end{aligned} \tag{4.8}$$

We further introduce a binary variable $\theta_\pi \in \{0,1\}$ for each sample $\pi \in [\Pi_1]$. When $\theta_\pi = 0$, it indicates that $\mathbb{I}(\mathbf{\Omega}^2_\pi$ satisfies (4.4) $| \, \mathbf{\Omega}^1) = 1$, i.e., all the constraints in the second-stage model (4.2) are satisfied; when $\theta_\pi = 1$, it indicates $\mathbb{I}(\mathbf{\Omega}^2_\pi$ satisfies (4.4) $| \, \mathbf{\Omega}^1) = 0$, i.e., all the constraints in the second-stage model (4.2) are not satisfied. Therefore, we can reformulate (4.8) as the following mixed-integer quadratic program:

$$
\begin{aligned}
\min_{\mathbf{\Omega}^1,\, \mathbf{\Omega}^2_\pi, \forall \pi \in [\Pi_1]} \quad & C_1\left(\mathbf{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\mathbf{\Omega}^2_\pi\right) && \text{(4.9a)} \\
\text{s.t.} \quad & (4.1\text{b}) - (4.1\text{g}), && \text{(4.9b)} \\
& (4.2\text{b})_\pi - (4.2\text{o})_\pi,\ \forall \pi \in [\Pi_1], && \text{(4.9c)} \\
& LS^t_{1\pi mn} \le \theta_\pi M_\pi,\ LS^t_{2\pi mn} \le \theta_\pi M_\pi, \forall (m,n) \in \mathcal{E}, \forall \pi \in [\Pi_1], && \text{(4.9d)} \\
& \sum_{\pi=1}^{\Pi_1} \theta_\pi \le \Pi_1 \eta, && \text{(4.9e)} \\
& \theta_\pi \in \{0,1\},\ \forall \pi \in [\Pi_1], && \text{(4.9f)}
\end{aligned}
$$

where M_π is a sufficiently large number for any $\pi \in [\Pi_1]$. Specifically, with constraints (4.9d) and (4.9e), we ensure that all but a few number (i.e., $\eta\Pi_1$) of samples in $[\Pi_1]$ satisfy the constraints in the second-stage model (4.2). That is, with the probability of $1-\eta$, the constraints in the second-stage model (4.2) are satisfied. In addition, as in the second-stage model (4.2), all constraints in (4.9) with a superscript t hold for all $t \in [T]$.

All of the constraints in (4.9) are convex except for $(4.2\text{f})_\pi, \pi \in [\Pi_1]$. Specifically, constraints (4.9b) are from the first-stage model, and all of them are linear constraints. Constraints (4.9d) - (4.9f) are also linear constraints. Constraints (4.9c), i.e., $(4.2\text{b})_\pi - (4.2\text{o})_\pi,\ \forall \pi \in [\Pi_1]$, are from the second-stage model, and they are either linear or second-order conic (SOC) constraints when the first-stage decision variables are given. However, as problem (4.9) needs to optimize both the first-stage and second-stage decision variables, constraints $(4.2\text{f})_\pi, \pi \in [\Pi_1]$ include bilinear terms ($z_{kn}u_{kl}$ and $w_{rn}f^t_{\pi r}$), by which these constraints are nonconvex. Specifically, after substituting x_k with (4.1d), we achieve bilinear terms $z_{kn}u_{kl}$ for $k \in [K], n \in \mathcal{N}, l \in [L]$. The bilinear terms can be linearized using McCormick inequalities.

McCormick inequalities are commonly used to linearize a bilinear term, say $w = xy$ with $x^L \leq x \leq x^U$ and $y^L \leq y \leq y^U$, in (mixed-integer) nonlinear programming (McCormick 1976). The general form of McCormick inequalities for $w = xy$ can be written as:

$$w \geq x^L y + x y^L - x^L y^L, \ w \geq x^U y + x y^U - x^U y^U,$$
$$w \leq x^U y + x y^L - x^U y^L, \ w \leq x y^U + x^L y - x^L y^U.$$

When x and y are both continuous variables, the above four McCormick inequalities provide convex and concave envelopes of the bilinear term xy. When at least one of x and y is binary, $w = xy$ can be implied by the above four McCormick inequalities, resulting in an equivalent mixed-integer linear reformulation of the bilinear expression.

Thus, for all $k \in [K], n \in \mathcal{N}, l \in [L]$, we replace $z_{kn}u_{kl}$ with β_{knl}, and add the following constraints

$$\beta_{knl} \geq 0, \ \beta_{knl} \geq z_{kn} + u_{kl} - 1,$$
$$\beta_{knl} \leq u_{kl}, \ \text{and } \beta_{knl} \leq z_{kn},$$

to enforce $\beta_{knl} = z_{kn}u_{kl}$, where both z_{kn} and u_{kl} are binary. Using the same technique, for all $r \in [R], n \in \mathcal{N}, t \in [T], \pi \in [\Pi_1]$, we denote $\hat{f}^t_{\pi rn} := w_{rn} f^t_{\pi r}$, where w_{rn} is binary, and add the following constraints:

$$\hat{f}^t_{\pi rn} \geq w_{rn} \underline{f}^t_{\pi r}, \ \hat{f}^t_{\pi rn} \geq f^t_{\pi r} + w_{rn} \overline{f}^t_{\pi r} - \overline{f}^t_{\pi r},$$
$$\hat{f}^t_{\pi rn} \leq f^t_{\pi r} + w_{rn} \underline{f}^t_{\pi r} - \underline{f}^t_{\pi r}, \ \text{and } \hat{f}^t_{\pi rn} \leq w_{rn} \overline{f}^t_{\pi r},$$

where $\overline{f}^t_{\pi r}$ and $\underline{f}^t_{\pi r}$ are the lower and upper bounds of $f^t_{\pi r}$. The bilinear term $w_{rn} g^t_{\pi r}$ can be managed similarly. As a result, the SAA formulation (4.9) is transformed into a mixed-integer convex quadratic program.

4.3.2 The PSAA Formulation

The SAA is relatively accurate when there are sufficient samples. However, more samples lead to more binary auxiliary variables (i.e., θ_π), greatly increasing the computational burden. Thus, we use partial sampling to reduce the computational difficulty and improve the solution quality. We extend the preliminary studies on partial sampling in Cheng et al. (2019) to approximate our proposed TCC model, which is more complicated than the single-stage chance-constrained model considered in Cheng et al. (2019), as evidenced in Liu et al. (2016). This leads to an extended PSAA model, referred to as *the PSAA model* for simplicity. It samples a part of the random parameters and estimates the probability distribution of the remainder. We first present the basic PSAA idea and then detail our PSAA model.

We consider a general chance constraint

$$\mathbb{P}\{g(\boldsymbol{x}, \boldsymbol{\xi}) \geq 0\} \geq 1 - \eta, \tag{4.10}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ and $\boldsymbol{\xi}_1$ is independent of $\boldsymbol{\xi}_2$. Clearly, $\mathbb{P}\{g(\boldsymbol{x}, \boldsymbol{\xi}) \geq 0\} = \mathbb{E}[\mathbb{I}(g(\boldsymbol{x}, \boldsymbol{\xi}) \geq 0)] = \mathbb{E}_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2}[\mathbb{I}(g(\boldsymbol{x}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \geq 0)] = \mathbb{E}_{\boldsymbol{\xi}_1}\mathbb{E}_{\boldsymbol{\xi}_2}[\mathbb{I}(g(\boldsymbol{x}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \geq 0)]$, where the third equation is because of the independence between $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. The PSAA idea then reformulates one of the above two expectations (i.e., $\mathbb{E}_{\boldsymbol{\xi}_1}$ and $\mathbb{E}_{\boldsymbol{\xi}_2}$) by its sample mean. For instance, if we replace the inner expectation $\mathbb{E}_{\boldsymbol{\xi}_2}$ by a sample mean of N independent samples of $\boldsymbol{\xi}_2$ (denoted by $\hat{\boldsymbol{\xi}}_2^1, \ldots, \hat{\boldsymbol{\xi}}_2^N$), then the PSAA formulation of (4.10) is as follows:

$$\begin{aligned}
&\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}_{\boldsymbol{\xi}_1}[\mathbb{I}(g(\boldsymbol{x}, \boldsymbol{\xi}_1, \hat{\boldsymbol{\xi}}_2^k))] \\
=&\frac{1}{N}\sum_{k=1}^{N}\mathbb{P}\{g(\boldsymbol{x}, \boldsymbol{\xi}_1, \hat{\boldsymbol{\xi}}_2^k) \geq 0\} \geq 1 - \eta,
\end{aligned}$$

which is further equivalent to the following

$$\mathbb{P}\{g(\boldsymbol{x}, \boldsymbol{\xi}_1, \hat{\boldsymbol{\xi}}_2^k) \geq 0\} \geq y_k, \forall k \in [N], \tag{4.11}$$

$$\frac{\sum_{k=1}^{N} y_k}{N} \geq 1 - \eta,\ y_k \geq 0, \forall k \in [N]. \tag{4.12}$$

In contrast to constraints (4.9d) - (4.9f) (SAA formulation), constraints (4.11)-(4.12) (PSAA formulation) introduce only continuous variable y_k, while N new chance constraints are added. In our proposed PSAA model, We will show that (4.11) has a convex approximation for *the planning problem* in this chapter, which contributes to the existing literature.

To apply the above PSAA idea, we need to have the sampled random parameters independent of the unsampled ones. We first convert the random vector $\boldsymbol{\xi}$ into an uncorrelated random vector $\boldsymbol{\xi}'$ using an affine transformation, thereby approximating the independence requirement. Specifically, let Σ be the covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\mu}$ be the mean vector of $\boldsymbol{\xi}$. Suppose that $\Sigma = U\Lambda U^\top$ is an eigenvalue decomposition of Σ, where U is an orthogonal matrix and Λ is a diagonal matrix with the eigenvalues of Σ on the diagonal. Without loss of generality, we assume that Λ_{11} is the largest eigenvalue of Σ. Let

$$\boldsymbol{\xi}' = \Lambda^{-\frac{1}{2}} U^\top (\boldsymbol{\xi} - \boldsymbol{\mu}), \quad \text{or equivalently,} \quad \boldsymbol{\xi} = U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}' + \mu.$$

It is straightforward to see that $\boldsymbol{\xi}'$ is an uncorrelated random vector with a mean of 0.

We partition $\boldsymbol{\xi}'$ as $(\xi_1', \boldsymbol{\xi}_2')$, where ξ_1' is the first component of $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}_2'$ is the vector of the other components. Note that when $\boldsymbol{\xi}$ is sampled, ξ_1' is the first principal component. The PSAA then considers Π_2 Monte Carlo samples $\boldsymbol{\xi}_{2\kappa}', \kappa \in [\Pi_2]$ of $\boldsymbol{\xi}_2'$ and approximates the probability of an event E as

$$\mathbb{P}_{(\xi_1', \boldsymbol{\xi}_2')}(E) \approx \frac{1}{\Pi_2} \sum_{\kappa=1}^{\Pi_2} \mathbb{P}_{\xi_1'}\left(E \,|\, \boldsymbol{\xi}_{2\kappa}'\right).$$

In our PSAA model, we retain the objective function of (4.8) and constraints (4.1b) – (4.1g) and $(4.2b)_\pi$ – $(4.2o)_\pi$. We also develop a different approximation of the chance constraint using the PSAA. To be compatible with the PSAA framework, given the first-stage variables $\mathbf{\Omega}^1$, instead of requiring (4.2b) – (4.2o) and (4.4) to be satisfied with a high probability by a specified $\mathbf{\Omega}^2$, we relax the chance constraint (4.6) to require the consistency of (4.2b) – (4.2o) and (4.4). That is, we consider the following chance constraint:

$$\mathbb{P}\left(\exists \text{ a solution satisfying (4.2b)} - \text{(4.2o), (4.4)} \,|\, \mathbf{\Omega}^1\right) \geq 1 - \eta. \tag{4.13}$$

Let $(4.2b)_\kappa$ – $(4.2o)_\kappa$ and $(4.4)_\kappa$ be a copy of (4.2b) – (4.2o) and (4.4), with $\boldsymbol{\xi}$ replaced by $(\xi_1', \boldsymbol{\xi}_{2\kappa}')$ for any $\kappa \in [\Pi_2]$. The PSAA approximates (4.13) with

$$\frac{1}{\Pi_2}\sum_{\kappa=1}^{\Pi_2} \mathbb{P}_{\xi_1'}\left(\xi_1' \in A\left(\mathbf{\Omega}^1, \kappa\right)\right) \geq 1 - \eta,$$

where $A(\mathbf{\Omega}^1, \kappa) := \{\xi_1' \,|\, (4.2b)_\kappa\text{-}(4.2o)_\kappa,\ (4.4)_\kappa \text{ are consistent with } \mathbf{\Omega}^1\}$. For a given $\mathbf{\Omega}^1$ and κ, $A(\mathbf{\Omega}^1, \kappa)$ is a convex set of ξ_1'. That is, if $\hat{\xi}_1' < \bar{\xi}_1'$ are both in $A(\mathbf{\Omega}^1, \kappa)$, then $\xi_1' \in A(\mathbf{\Omega}^1, \kappa)$ for all $\xi_1' \in [\hat{\xi}_1', \bar{\xi}_1']$. Let $\Psi(\cdot)$ be the cumulative distribution function (CDF) of ξ_1'. Then,

$$\mathbb{P}_{\xi_1'}\left(\xi_1' \in A\left(\mathbf{\Omega}^1, \kappa\right)\right) = \sup_{Z_1, Z_2}\left\{\Psi(Z_2) - \Psi(Z_1) \mid Z_1, Z_2 \in A\left(\mathbf{\Omega}^1, \kappa\right)\right\}.$$

Therefore, our PSAA model is as follows.

$$\begin{aligned}
\min\ & C_1\left(\mathbf{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\mathbf{\Omega}_\pi^2\right) \\
\text{s.t. } & (4.1b) - (4.1g), && (4.14a)\\
& (4.2b)_\pi - (4.2o)_\pi,\ \forall\, \pi \in [\Pi_1], && (4.14b)\\
& (\mathbf{\Omega}^1, \mathbf{\Omega}_{1\kappa}^2) \text{ satisfies } (4.2b)_\kappa - (4.2o)_\kappa,\ (4.4)_\kappa \text{ with } (Z_{1\kappa}, \boldsymbol{\xi}_{2\kappa}'),\ \forall\, \kappa \in [\Pi_2], && (4.14c)
\end{aligned}$$

$$(\boldsymbol{\Omega}^1, \boldsymbol{\Omega}^2_{2\kappa}) \text{ satisfies } (4.2\text{b})_\kappa - (4.2\text{o})_\kappa,\ (4.4)_\kappa \text{ with } (Z_{2\kappa}, \boldsymbol{\xi}'_{2\kappa}),\ \forall \kappa \in [\Pi_2], \tag{4.14d}$$

$$\Psi(Z_{2\kappa}) - \Psi(Z_{1\kappa}) \geq \eta_\kappa,\ \forall \kappa \in [\Pi_2], \tag{4.14e}$$

$$\sum_{\kappa=1}^{\Pi_2} \eta_\kappa \geq \Pi_2(1-\eta), \tag{4.14f}$$

$$\eta_\kappa \geq 0,\ \forall \kappa \in [\Pi_2]. \tag{4.14g}$$

Here, the decision variables are $\boldsymbol{\Omega}^1$, $\boldsymbol{\Omega}^2_\pi$ for any $\pi \in [\Pi_1]$, and $\boldsymbol{\Omega}^2_{1\kappa}, \boldsymbol{\Omega}^2_{2\kappa}, Z_{1\kappa}, Z_{2\kappa}, \eta_\kappa$ for any $\kappa \in [\Pi_2]$.

For the model to be practical, we need to estimate the CDF $\Psi(\cdot)$ of ξ'_1. Two types of methods are primarily used for estimating distributions: parametric and nonparametric estimation methods. Parametric estimation methods assume that the sample data conform to a parametrized family of probability distributions, and the sample data are used to find the best-fitting parameters. In contrast, nonparametric estimation methods do not depend on any prior assumption of the distribution family, and they fit the distribution according to the characteristics and properties of the data. Here we make no assumptions on the distribution of ξ'_1 and estimate $\Psi(\cdot)$ using the kernel density estimation, a commonly used nonparametric method proposed by Rosenblatt (Rosenblatt 1956) and Parzen (Parzen 1962).

Let Π_3 be the total number of samples of ξ'_1, and $\xi'_{1\tau}$ be a specific sample for any $\tau \in [\Pi_3]$. The kernel density estimation of the probability density function ψ of ξ'_1 can be written as

$$\psi(\xi'_1) \approx \frac{1}{\Pi_3 h} \sum_{\tau=1}^{\Pi_3} \phi\left(\frac{\xi'_1 - \xi'_{1\tau}}{h}\right),$$

where h is a user-specified bandwidth parameter and ϕ is a kernel function. Among the popular choices, we choose the standard normal density function as the kernel function for our estimation. Thus, the CDF of ξ'_1 can be estimated by

$$\Psi(\xi'_1) \approx \frac{1}{\Pi_3} \sum_{\tau=1}^{\Pi_3} \Phi\left(\frac{\xi'_1 - \xi'_{1\tau}}{h}\right), \tag{4.15}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

We further approximate $\Phi(\cdot)$ in (4.15) using the following piecewise linear function:

$$\Phi(x) \approx \begin{cases} \min_{\epsilon\in[\Delta_1]} \{a_\epsilon x + \alpha_\epsilon\} & \text{if } x \geq 0.5 \\ \max_{\zeta\in[\Delta_2]} \{a_\zeta x + \alpha_\zeta\} & \text{if } x < 0.5, \end{cases} \tag{4.16}$$

where Δ_1 and Δ_2 are the numbers of pieces used to approximate the upper half and the lower half of $\Phi(\cdot)$, respectively. An example of such an approximation is depicted in the following Fig. 4.2.

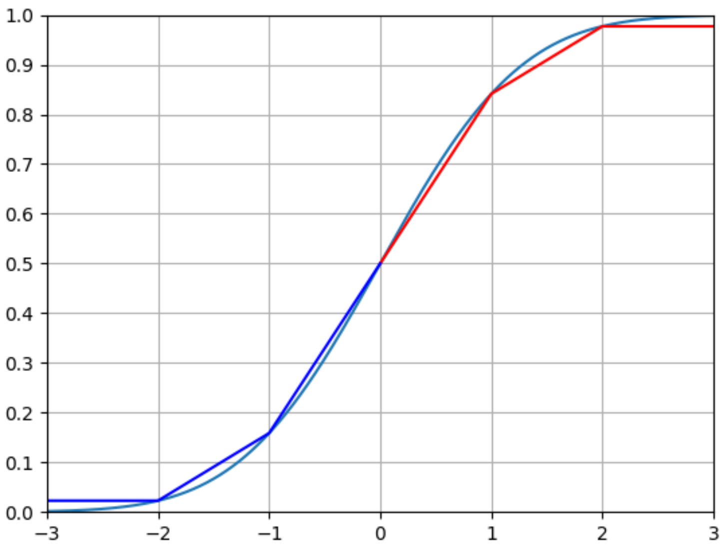

Figure 4.2: Piecewise linear approximation of the CDF of the standard normal distribution $\Phi(\cdot)$ ($\Delta_1 = \Delta_2 = 3$)

When the probability level $1-\eta$ is close to 1, $\Phi(\frac{Z_{2\kappa}-\xi'_{1\tau}}{h})$ is usually greater than 0.5 and thus concave, while $\Phi(\frac{Z_{1\kappa}-\xi'_{1\tau}}{h})$ is usually less than 0.5 and thus convex(Karimi et al. 2021). Therefore, with the approximation in (4.16), we can remove the min and max operators in (4.16) and approximate constraint (4.14e) by the following constraints:

$$\rho^\kappa_{1\tau} \geq a_\zeta \left(\frac{Z_{1\kappa} - \xi'_{1\tau}}{h}\right) + \alpha_\zeta,\ \forall\, \zeta \in [\Delta_2],\, \tau \in [\Pi_3],\, \kappa \in [\Pi_2],$$

$$\rho_{2\tau}^{\kappa} \le a_{\epsilon}\left(\frac{Z_{2\kappa}-\xi'_{1\tau}}{h}\right)+\alpha_{\epsilon},\ \forall\,\epsilon\in[\Delta_1],\,\tau\in[\Pi_3],\,\kappa\in[\Pi_2],$$

$$\frac{1}{\Pi_3}\sum_{\tau=1}^{\Pi_3}(\rho_{2\tau}^{\kappa}-\rho_{1\tau}^{\kappa})\ge\eta_{\kappa},\ \forall\,\kappa\in[\Pi_2].$$

Finally, the bilinear terms in (4.14b) – (4.14d) can be linearized using McCormick inequalities, as in the SAA model. As a result, the PSAA model (4.14) is simplified to a mixed-integer convex quadratic program.

4.4 Numerical Results

We conduct two sets of experiments on the IEEE 33-Bus system and the IEEE 123-Bus system using real data acquired from Pecan Street Inc. and ERCOT. We first compare the effectiveness of the SAA and PSAA models and then investigate the potential benefits of installing ES units. We use in-sample and out-of-sample tests to validate the quality of the obtained solutions to the planning problem. All numerical tests are executed on the high-performance computing (HPC) cluster of Ieria with 27 computing nodes. We allocate four CPUs to every instance, and every CPU is allocated 4 GB of memory. CPLEX 22.1.0, with its default setting, is used to solve all optimization models. For ease of exposition, we use the following flowchart in Fig. 4.3 to summarize the sequential steps we follow to perform the numerical experiments in this section.

4.4.1 IEEE 33-Bus System

We first consider the modified IEEE 33-Bus radial distribution network examined in Fathabad et al. (2020) (see Fig. 4.4). In the network, Bus 0 is connected to the major transmission network, from which we can purchase active and reactive power via Bus 0 if needed. Buses 1–32 are connected to Bus 0, directly or indirectly. Two DDG units are located at Buses 15 and 29, and three reactive power sources are located at Buses

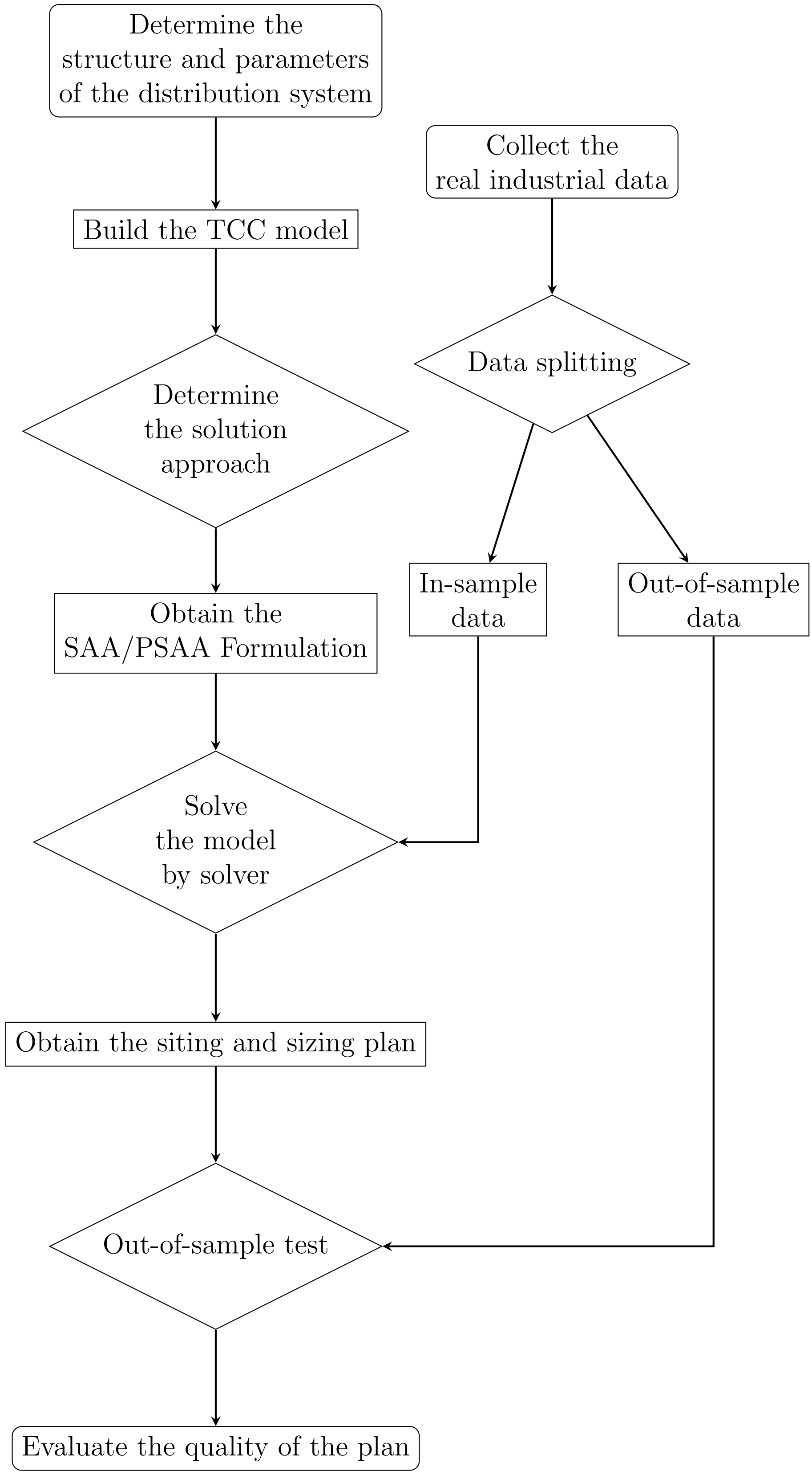

Figure 4.3: The Procedure of Numerical Experiments

11, 13, and 32. The reactive power sources are of the hybrid (capacitive and inductive) compensator type, and they can both generate and absorb reactive power to stabilize the voltage. Nevertheless, our model can also consider other types. With the fixed location of the DDG units and reactive power sources, we then focus on the location and capacity planning of candidate RDG and ES units to be installed.

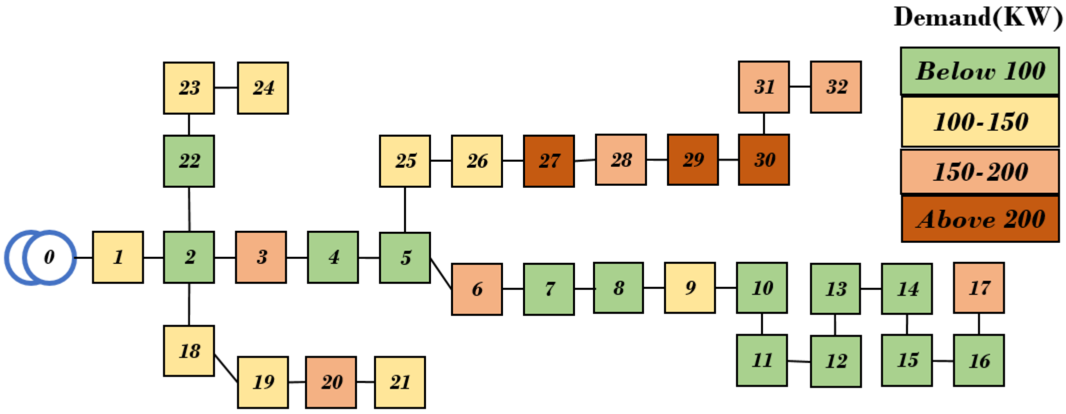

Figure 4.4: IEEE 33-Bus Distribution Network

4.4.1.1 Data

We consider two sources of uncertainties: the weekly active/reactive load at each bus, i.e., d^t_{pn}/d^t_{qn}, and the renewable generation efficiency of each candidate RDG unit, i.e., $s^t_k \in [0,1]$. The active-load data are obtained from Pecan Street Inc. and the wind generation data from ERCOT. The reactive-load data are randomly generated based on the bounds of the total reactive-power output. Specifically, the reactive-power load is uniformly generated in the interval $[-0.01, 0.019]$. Here $[-0.01, 0.019]$ is an interval, which specifies a range of possible values of reactive-power load that we can possibly generate. By using such a random data generation, we obtain different values of reactive-power loads at different buses in different periods. All data are for a range of 4 years, leading to 208 ($=$ 52 weeks $\times$ 4) data samples for each specific random variable. To perform more practical tests, we randomly generate more data samples to better demonstrate our proposed models' effectiveness, via in-sample and out-of-sample tests. To that end, we first calculate the mean value and covariance matrix using the given data samples, and then generate 3792 data samples by following the multivariate log-normal distribution, which has been widely adopted in academia and industry in similar scenarios (Park et al. 2015, Qin et al. 2012, Qian et al. 2010). Thus, we have 4000 data samples in total.

4.4.1.2 Parameters

All of the parameters used in our experiments are slightly modified based on the parameters[1] used in Fathabad et al. (2020). For instance, the electricity price of purchasing active/reactive power from the main grid is mainly from ERCOT. The detailed modification is as follows. In the first stage, the setup costs of candidate RDG units c^0_{kn} are uniformly generated in the interval $[0.95 \times 2000, 1.05 \times 2000]$. The size-based investment costs c^1_k and maintenance costs c^2_k of candidate RDG units are uniformly generated in the intervals $[0.9 \times 238, 1.1 \times 238]$ and $[0.9 \times 4, 1.1 \times 4]$, respectively. The setup costs of candidate ES units d^0_{rn} are uniformly generated in the interval $[0.9 \times 200, 1.1 \times 200]$. The size-based investment costs d^1_r and maintenance costs d^2_r of candidate ES units are both uniformly generated in the interval $[0.9 \times 2, 1.1 \times 2]$. The active power purchase prices c^t_p are uniformly generated in the interval $[0.9 \times 130, 1.1 \times 130]$. The reactive power purchase prices c^t_q are uniformly generated in the interval $[0.9 \times 4, 1.1 \times 4]$. The emission costs for the DDG units at Buses 15 and 29 are $c^f_{15} = c^f_{29} = 630$. The emission factor ω of the DDG units is 3 kg/MWh. A maximum of $\bar{K} = 3$ out of $K = 4$ candidate RDG units are to be installed in this distribution network. The maximum number of ES units to be installed is $R = 3$. The active-power output bounds $(\underline{p}^t_n, \bar{p}^t_n)$ of both DDG units are $(0.5, 4.5)$. The reactive-power output bounds $(\underline{q}^t_n, \bar{q}^t_n)$ are $(-0.1, 0.2)$, $(-0.15, 0.25)$, and $(-0.1, 0.2)$ for the three reactive-power sources at Buses 11, 13, and 32, respectively. We further consider four types of RDG units ($\bar{x}_1 = 4$ MW, $\bar{x}_2 = 5$ MW, $\bar{x}_3 = 6$ MW, and $\bar{x}_4 = 7$ MW). The maximum capacity $\bar{y}_r$ of an ES unit is 3 MW, and the minimum capacity $\underline{y}_r$ is 0. The initial power level b^0_r of a candidate ES unit is set to 0. When an ES

[1]See https://www.dropbox.com/s/psqv9yr3atg46bk. Accessed: Jul. 2022.

unit is charged, the unit cost e_1 is 0.1, whereas the discharging cost e_2 is 0.1. The energy loss factor γ is set to 0.9.

4.4.2 IEEE 123-Bus System

We then consider the commonly used IEEE 123-Bus radial distribution network (Bazrafshan and Gatsis 2017, Daratha et al. 2013, Zhang et al. 2013) (see Fig. 4.5). In the network, Bus 149 is connected to the major transmission network. Eight DDG units are located at Buses 8, 25, 44, 57, 67, 87, 97, and 108, and twelve reactive power sources are located at Buses 7, 14, 15, 25, 47, 54, 62, 68, 80, 91, 98, and 109.

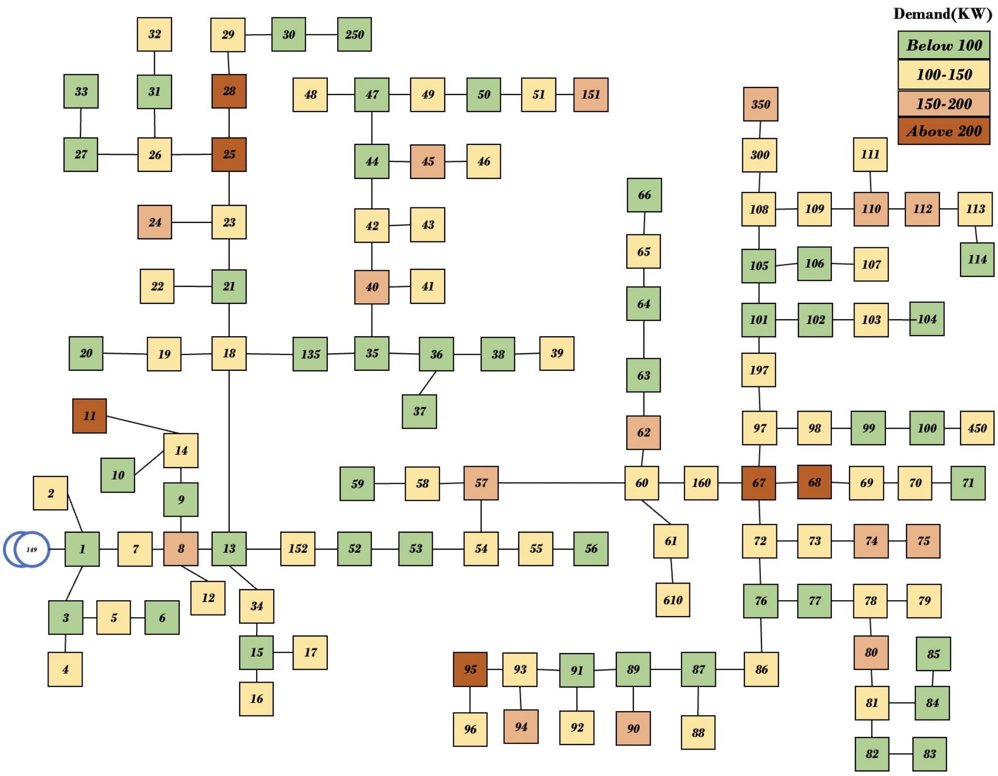

Figure 4.5: IEEE 123-Bus Distribution Network

4.4.2.1 Data

All the data are obtained following the same process used for the IEEE 33-Bus system. The only difference is the dimension of uncertainty. Here we consider the renewable generation efficiency of each candidate RDG unit s_k^t uncertain and use an estimated value for the weekly active/reactive load at each bus, i.e., d_{pn}^t/d_{qn}^t. In particular, as the IEEE 123-Bus system is of large scale, both the SAA and PSAA formulation become difficult to solve when the test system is large. Thus, to better show the performance of these two approaches, we consider the system loads are given and the renewable generation is uncertain, by which the computational difficulty is relatively reduced.

4.4.2.2 Parameters

We continue to use the parameters designed for the IEEE 33-Bus system, except that some parameters are modified as follows. First, the electrical resistance ($\mathfrak{R}_{mn}$) and reactance ($\mathfrak{X}_{mn}$) of each line $(m, n) \in \mathcal{E}$ and the upper/lower bound of voltage magnitude (i.e., $\underline{v}/\bar{v}$) at each bus are obtained from the IEEE PES Test Feeders[2]. Second, we consider a maximum of $\bar{K} = 6$ out of $K = 8$ candidate RDG units to be installed in this distribution network. We consider three types of RDG units ($\bar{x}_1 = 8$ MW, $\bar{x}_2 = 10$ MW, and $\bar{x}_3 = 12$ MW). The maximum capacity $\bar{y}_r$ of an ES unit is 6 MW. The maximum number of candidate ES units to be installed is $R = 6$. The reactive-power output bounds $(\underline{q}_n^t, \bar{q}_n^t)$ are (-0.15, 0.25).

4.4.3 Decomposition Framework

To reduce the computational difficulty of solving both the SAA and PSAA formulations, we adopt the Benders decomposition algorithm (Benders 1962) to improve the computational efficiency. Specifically, we first linearize the SOCP constraints (4.2l) and (4.2m) as Fathabad et al. (2020) does by using the polyhedral ϵ-approximation in Ben-Tal and Nemirovski (2001). With such an approximation, both the SAA formulation (4.9) and PSAA formulation (4.14) are transformed into mixed-integer linear programming (MILP) formulations, which can be used practically in large-scale settings.

[2]See https://cmte.ieee.org/pes-testfeeders/resources/. Accessed: Jul. 2022.

For each MILP formulation, we then decompose the problem into two parts: a master problem and a set of subproblems. The master problem includes all the integer variables and associated constraints, and the subproblems contain the remaining continuous variables and associated constraints. As such, we iteratively solve the master problem and subproblems until convergence. At each iteration, the master problem is solved to optimality, and its optimal solution is then used to construct the subproblems. Feasible and optimality cuts are generated after solving the subproblems and added back to the master problem.

Note that we have Π_1 subproblems for the SAA formulation, where each sample $\pi \in [\Pi_1]$ corresponds to one subproblem, as shown in (4.9c), and these subproblems can be solved in parallel. For the PSAA formulation, we have $\Pi_1 + 1$ subproblems, where each sample $\pi \in [\Pi_1]$ corresponds to one subproblem, as shown in (4.14b), and constraints (4.14c) – (4.14g) are included in one subproblem.

4.4.4 SAA vs. PSAA

Here we analyze the performance of the SAA and PSAA models using the data and parameters mentioned above. We thus ignore ES units and consider $T = 1$. Specifically, terms related to ES units, including w_{rn}, y_r, f_r^t, g_r^t, and b_r^t, are temporarily removed from the models.

We first divide the 4000 data samples into two sets for the experiment: a training data set and a testing data set. The former is used to obtain our planning decision in the first stage, and the latter is used to test the effectiveness of the obtained decision. To make full use of the real data and better simulate real-world decision-making, we ensure that the training data are selected from the first 208 samples, as it is very straightforward to feed the available historical real data into an optimization model to support decision-making. Each data sample is used as a scenario in the SAA and PSAA models. We solve the SAA and PSAA models using the training data and obtain two optimal sizing/siting plans.

We then compare the performance of the plans using the testing data. Specifically, we calculate the first-stage cost and the average second-stage cost of all test samples for each plan. To verify our approximation formulations and demonstrate their ability to ensure that the demand can be satisfied with a high probability, we also calculate the actual feasible probability of the test samples. This probability is defined as the percentage of the test samples for which constraints (4.2b) – (4.2o) and (4.4) can be simultaneously satisfied.

The experiments are conducted using different settings for (i) the size of the training data (i.e., $\Pi_1 = \Pi_2$), (ii) the maximum running time (i.e., time limit) of the solver (denoted by ϑ in hours), and (iii) the desired feasible probability of the chance constraint (i.e., $1-\eta$), as shown in Tables 4.1 – 4.4. The training data size $\Pi_1 = \Pi_2 \in \{60, 100, 140\}$ for the IEEE 33-Bus system and $\Pi_1 = \Pi_2 \in \{30, 45, 60\}$ for the IEEE 123-Bus system. The maximum running time $\vartheta \in \{4, 8, 12\}$ for the IEEE 33-Bus system and $\vartheta \in \{10, 13, 16\}$ for the IEEE 123-Bus system. The desired feasible probability $1 - \eta \in \{0.8, 0.9\}$ for both systems. These settings lead to $18 = 3 \times 3 \times 2$ combinations in total for each system. For some instances, when both models are too large to be solved to optimality within the given time limits, we take the incumbent solutions returned by the solver as the optimal solutions. We also record the relative optimality gap, $(z_p - z_d)/z_p$, where z_p is the primal objective bound (i.e., the incumbent objective value) and z_d is the dual objective bound (i.e., the lower bound for minimization problems). Intuitively, a smaller gap indicates a better-quality incumbent solution.

We illustrate the performance of both models in Tables 4.1 – 4.4. The columns in the SAA/PSAA section represent, from left to right, the first-stage cost, the average second-stage cost of testing samples, the average total cost, the relative optimality gap, and the actual feasible probability.

Compared with the SAA model, the PSAA model leads to lower first-stage costs, lower second-stage costs, and thus lower total costs in all cases. The lower costs indicate that the siting and sizing decisions provided by the PSAA approach help more effectively satisfy the same required demands than those provided by the SAA approach. Thus, compared to the SAA solutions, the PSAA solutions require fewer RDG units to be installed and/or the installed RDG units can be of smaller capacity. Moreover, the effective plans made by the PSAA approach lead to lower operating costs for the distribution grid than the plans made by the SAA approach. A direct comparison of the total costs is shown in Fig. 4.6, where the horizontal axis represents the training-data size, the vertical axis represents the total cost, and different colors represent different running times. The total costs of the SAA solutions are labeled with triangular symbols, and the total costs of the PSAA solutions are marked with circular symbols.

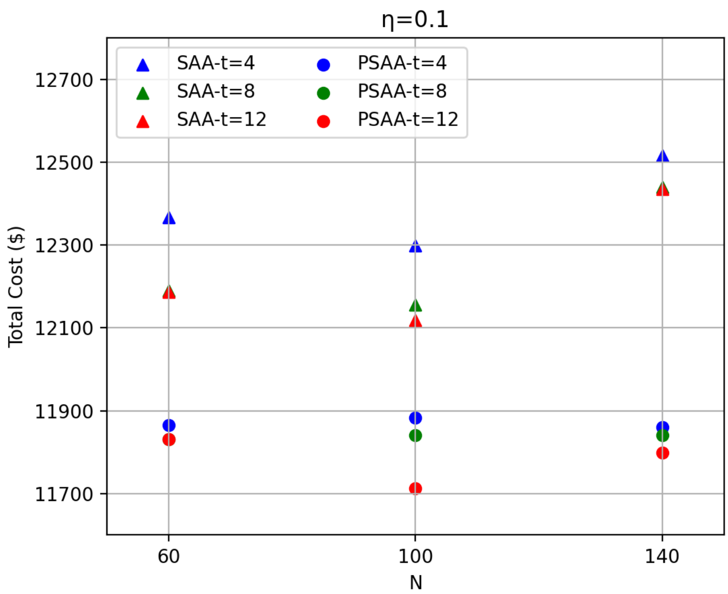

Figure 4.6: SAA vs. PSAA solutions in the total cost ($\eta = 0.1$)

Within the same time limits, we observe that the relative optimality gap of the PSAA solution is always less than that of the SAA solution in all cases. Specifically, for the cases where the PSAA model can solve the instance to optimality (i.e., the optimality gap is 0) within the time limits, we report the corresponding computational time in hours used by the PSAA model in the column "Gap" and label it by $\star$. The SAA model cannot solve any instance to optimality within the time limits. The result clearly indicates that the PSAA model is more computationally efficient than the SAA model. The difference is due to how the chance constraint is dealt with in the models. The SAA model introduces Π_1 binary variables (i.e., θ_π), whereas the PSAA model introduces Π_2 continuous variables (i.e., η_κ). Although there are more continuous variables and more constraints in the PSAA model, the binary variables in the SAA model are more difficult to manage. A direct comparison of the gap is shown in Fig. 4.7, where the vertical axis represents the relative optimality gap.

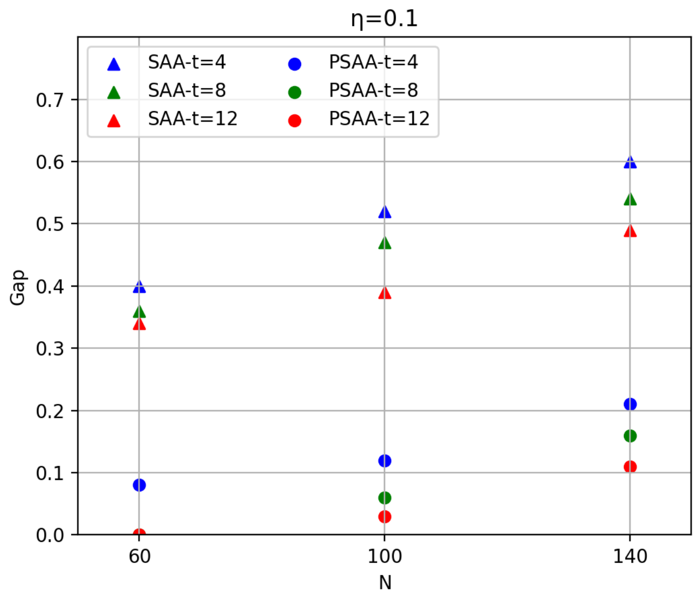

Figure 4.7: SAA vs. PSAA solutions in the optimality gap ($\eta = 0.1$)

To further illustrate the computational improvement of the PSAA approach compared with the SAA approach, we summarize the optimality gap improvement (i.e., reduction) from the SAA approach to the PSAA approach in Fig. 4.8 and Fig. 4.9. In the figures, the horizontal axis represents the setting of η and Π_1. For instance, "$0.1 - 60$" means that $\eta = 0.1$ and $\Pi_1 = 60$. The vertical axis represents the optimality gap improvement in percentage, as given by:

$$\frac{|\text{the gap by PSAA} - \text{the gap by SAA}|}{\text{the gap by SAA}} \times 100\%.$$

From Fig. 4.8 and Fig. 4.9, we find that the improvement is mostly above 50% and even reaches 100% when the PSAA solves an instance to optimality.

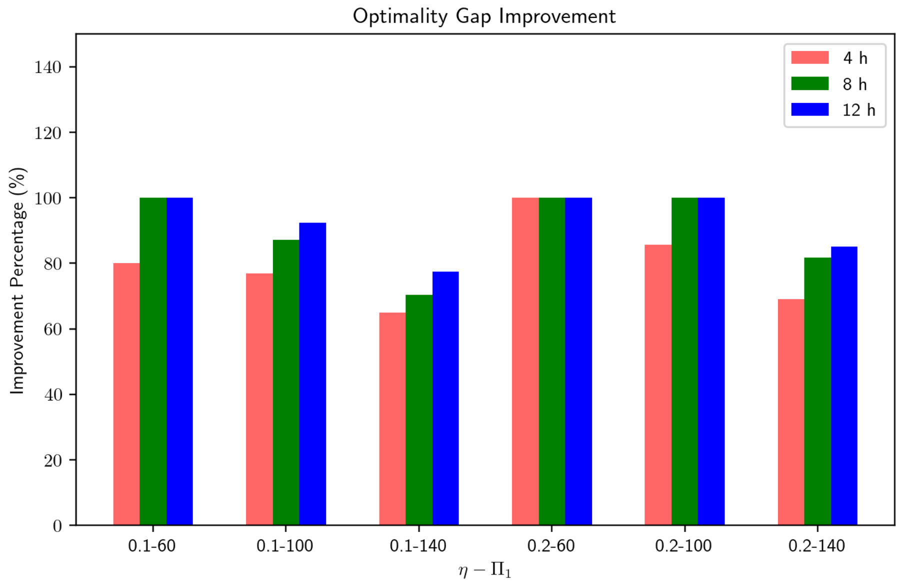

Figure 4.8: Optimality Gap Improvement (IEEE 33-Bus system)

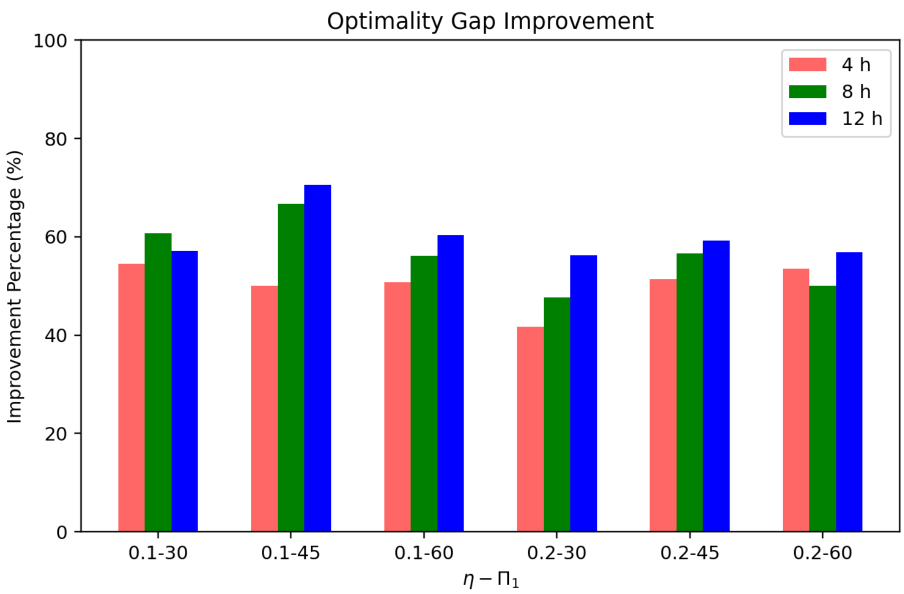

Figure 4.9: Optimality Gap Improvement (IEEE 123-Bus system)

In each of the tested cases, the actual feasible probability of the PSAA solution is clearly higher than that of the SAA solution and is almost equal to the desired probability.

This indicates that the PSAA performs better than the SAA as an approximation method for the chance constraint. In fact, the actual feasible probability of the PSAA solution is less than 1% different from the desired solution, which in practice will give the grid decision-makers more control of the confidence level. A direct comparison is shown in Fig. 4.10, where the vertical axis represents the actual feasible probability.

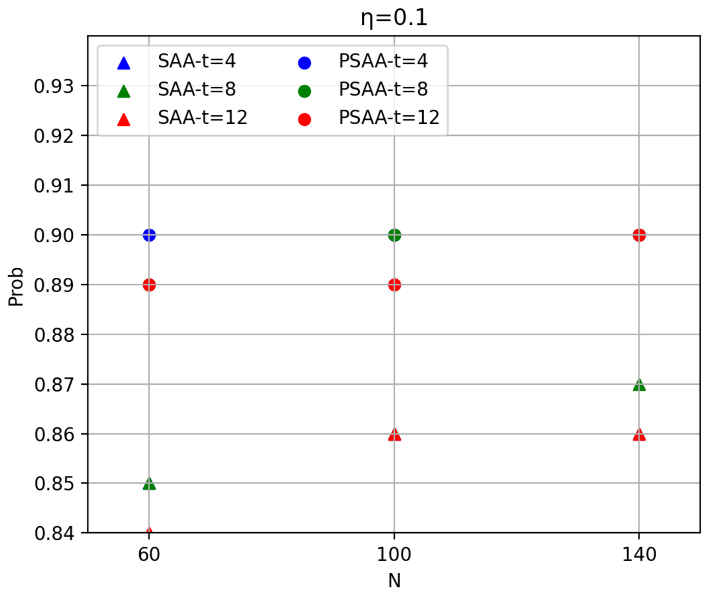

Figure 4.10: SAA vs. PSAA solutions in the actual feasible probability ($\eta = 0.1$)

To conclude this section, we illustrate the performance difference between the two models with a specific example. When $\Pi_1 = \Pi_2 = 140$, $\vartheta = 8$, and $\eta = 0.1$ in the IEEE 33-Bus system, the SAA solution sites (sizes) the candidate RDG units at Buses 3 (6 MW), 9 (6 MW), and 12 (4 MW), and the PSAA solution sits (sizes) the candidate RDG units at Buses 3 (5 MW), 8 (4 MW), and 26 (6 MW). Thus, compared to the SAA solution, the PSAA solution results in a lower first-stage cost, due to the smaller total capacities of the installed RDG units. In addition, the output p_n^t of the DDG unit at Bus 29 in the PSAA solution is significantly less than that in the SAA solution for most testing data, and the load-shedding penalties of the PSAA solution are also much less than those of the SAA solution. The two factors above account for most of the difference between the methods in the second-stage costs. We know that Buses 27, 29, and 30 have higher loads and that the DDG units have been installed at Buses 15 and 29. Thus, it is reasonable to place a candidate RDG unit at Bus 26 to reduce the output pressure on the DDG unit at Bus 29. In addition, in a distribution network structure, it is likely that the power purchased from Bus 0 and generated by the RDG unit at Bus 3 (6 MW) or Bus 3 (5 MW) is mainly used to fulfill the loads at Buses 1-4 and 18-24. Given the desired feasible probability of 0.9 (i.e., the power system operator expects to satisfy the load with a confidence level of 0.9), the PSAA solution shows that a 5 MW RDG unit is sufficient to satisfy the load, and thus a 6 MW unit (given by the SAA solution) may be excessive.

Table 4.1: IEEE 33-Bus System: SAA vs. PSAA ($\eta = 0.1$)

Π_1	$\vartheta\,(h)$	SAA					PSAA				
		1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
60	4	10674.5	1692.8	12367.3	0.40	0.85	10226.3	1639.2	11865.5	0.08	0.90
	8	10510.0	1679.4	12189.4	0.36	0.85	10184.6	1645.8	11830.4	7.33$\star$	0.89
	12	10510.0	1679.4	12189.4	0.34	0.84	10184.6	1645.8	11830.4	7.33$\star$	0.89
100	4	10586.3	1713.0	12299.3	0.52	0.86	10275.5	1607.6	11883.1	0.12	0.90
	8	10447.2	1708.5	12155.7	0.47	0.86	10224.5	1616.2	11840.7	0.06	0.90
	12	10397.4	1721.3	12118.7	0.39	0.86	10105.6	1606.5	11712.1	0.03	0.89
140	4	10800.2	1716.0	12516.2	0.60	0.86	10231.8	1629.0	11860.8	0.21	0.90
	8	10749.0	1691.2	12440.2	0.54	0.87	10209.0	1632.7	11841.7	0.16	0.90
	12	10723.9	1711.9	12435.8	0.49	0.86	10173.0	1625.1	11798.1	0.11	0.90

Table 4.2: IEEE 33-Bus System: SAA vs. PSAA ($\eta = 0.2$)

Π_1	$\vartheta\,(h)$	SAA					PSAA				
		1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
60	4	9947.9	1756.2	11704.1	0.21	0.78	9536.3	1723.3	11259.6	2.55$\star$	0.80
	8	9947.9	1756.2	11704.1	0.12	0.78	9536.3	1723.3	11259.6	2.55$\star$	0.80
	12	9664.2	1742.8	11407.0	0.08	0.77	9536.3	1723.3	11259.6	2.55$\star$	0.80
100	4	9772.3	1722.6	11494.9	0.35	0.78	9652.7	1720.1	11372.8	0.05	0.80
	8	9651.9	1743.9	11395.8	0.27	0.78	9574.9	1652.2	11227.1	4.73$\star$	0.80
	12	9633.8	1739.7	11373.5	0.21	0.77	9574.9	1652.2	11227.1	4.73$\star$	0.80
140	4	10039.0	1740.1	11779.1	0.42	0.79	9705.0	1692.1	11397.1	0.13	0.80
	8	9974.3	1748.7	11723.0	0.33	0.79	9639.5	1683.0	11322.5	0.06	0.80
	12	9974.3	1748.7	11723.0	0.27	0.79	9592.1	1666.8	11258.9	0.04	0.80

Table 4.3: IEEE 123-Bus System: SAA vs. PSAA ($\eta = 0.1$)

Π_1	$\vartheta\,(h)$	SAA					PSAA				
		1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
30	10	26715.3	6422.7	33138.0	0.33	0.86	25937.3	6396.0	32333.3	0.15	0.91
	13	26127.5	6248.7	32376.2	0.28	0.86	25372.6	6210.7	31583.3	0.11	0.90
	16	25972.4	6255.3	32227.7	0.21	0.85	25047.2	6251.8	31299.0	0.09	0.90
45	10	27019.2	6392.3	33411.5	0.42	0.87	26252.8	6305.3	32558.1	0.21	0.90
	13	26407.7	6268.9	32676.6	0.39	0.86	26043.7	6365.3	32409.0	0.13	0.90
	16	26149.3	6274.3	32423.6	0.34	0.86	25392.6	6216.6	31609.2	0.10	0.89
60	10	26931.0	6561.9	33492.9	0.61	0.87	25892.4	6410.8	32303.2	0.30	0.91
	13	26772.4	6347.2	33119.6	0.57	0.87	25465.3	6304.3	31769.6	0.25	0.91
	16	26073.5	6403.6	32477.1	0.53	0.87	25428.0	6309.8	31737.8	0.21	0.90

Table 4.4: IEEE 123-Bus System: SAA vs. PSAA ($\eta = 0.2$)

Π_1	$\vartheta\,(h)$	SAA					PSAA				
		1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
30	10	25894.3	6407.1	32301.4	0.24	0.78	25283.4	6392.2	31675.6	0.14	0.80
	13	25607.6	6379.2	31986.8	0.21	0.78	25076.9	6359.4	31436.3	0.11	0.80
	16	25313.7	6392.4	31706.1	0.16	0.77	24764.0	6271.2	31035.2	0.07	0.79
45	10	26021.9	6307.6	32329.5	0.37	0.78	25506.3	6293.2	31799.5	0.18	0.80
	13	25528.3	6517.4	32045.7	0.30	0.78	24892.6	6268.1	31160.7	0.13	0.80
	16	24986.1	6492.2	31478.3	0.27	0.78	24508.3	6238.9	30747.2	0.11	0.80
60	10	26328.0	6398.2	32726.2	0.56	0.78	26017.3	6344.5	32361.8	0.26	0.81
	13	26148.7	6362	32510.7	0.48	0.78	25370.4	6308.8	31679.2	0.24	0.80
	16	25693.3	6417.4	32110.7	0.44	0.78	25091.2	6284.5	31375.7	0.19	0.80

Table 4.5: IEEE 33-Bus System: Storage vs. No Storage ($\eta = 0.1$)

(ϱ_1, ϱ_2)	without energy storage					with energy storage				
	1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
(0.5,0.5)	10592.0	5243.3	15835.3	0.24	0.89	9952.0	5501.4	15453.4	0.26	0.90
(0.4,0.6)	10464.3	5372.8	15837.1	0.23	0.90	9873.8	5427.6	15301.4	0.29	0.91
(0.6,0.4)	10726.9	5194.2	15921.1	0.27	0.89	9908.2	5462.9	15371.1	0.28	0.90
(0.3,0.3)			*			10143.7	5576.1	15719.8	0.38	0.89
(0.2,0.4)			*			10471.5	5560.8	16032.3	0.39	0.89
(0.4,0.2)			*			10239.2	5602.5	15841.7	0.43	0.89

Table 4.6: IEEE 33-Bus System: Storage vs. No Storage ($\eta = 0.2$)

(ϱ_1, ϱ_2)	without energy storage					with energy storage				
	1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
(0.5,0.5)	10527.7	5293.7	15821.4	0.16	0.80	9903.2	5460.4	15363.6	0.22	0.82
(0.4,0.6)	10332.5	5268.3	15600.8	0.20	0.79	9764.0	5503.9	15267.9	0.26	0.80
(0.6,0.4)	10200.4	5372.9	15573.3	0.21	0.80	9672.4	5423.5	15095.9	0.23	0.81
(0.3,0.3)			*			10021.3	5690.3	15711.6	0.33	0.80
(0.2,0.4)			*			10206.5	5575.1	15781.6	0.40	0.80
(0.4,0.2)			*			10164.8	5625.0	15789.8	0.38	0.81

Table 4.7: IEEE 123-Bus System: Storage vs. No Storage ($\eta = 0.1$)

(ϱ_1, ϱ_2)	without energy storage					with energy storage				
	1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
(0.5,0.5)	28274.9	19707.3	47982.2	0.44	0.89	26922.5	20311.6	47234.1	0.46	0.90
(0.4,0.6)	27944.6	19638.2	47582.8	0.37	0.89	26327.3	20610.4	46937.7	0.46	0.91
(0.6,0.4)	27826.9	19648.2	47475.1	0.39	0.89	26281.7	20409.3	46691.0	0.49	0.90
(0.3,0.3)			*			28037.3	20416.3	48453.6	0.57	0.90
(0.2,0.4)			*			27614.0	20882.5	48496.5	0.54	0.89
(0.4,0.2)			*			27932.4	20741.3	48673.7	0.52	0.90

Table 4.8: IEEE 123-Bus System: Storage vs. No Storage ($\eta = 0.2$)

(ϱ_1, ϱ_2)	without energy storage					with energy storage				
	1st ($)	2nd ($)	Total cost ($)	Gap	Prob	1st ($)	2nd ($)	Total cost ($)	Gap	Prob
(0.5,0.5)	27970.4	19504.8	47475.2	0.38	0.80	26392.7	19903.2	46295.9	0.37	0.81
(0.4,0.6)	27822.6	19762.4	47585.0	0.29	0.80	27041.5	20113.4	47154.9	0.41	0.81
(0.6,0.4)	28203.0	19793.0	47996.0	0.33	0.80	26808.3	20513.6	47321.9	0.35	0.81
(0.3,0.3)	29409.4	19442.7	48852.1	0.52	0.78	27751.4	19862.4	47613.8	0.45	0.79
(0.2,0.4)	29143.6	19627.1	48770.7	0.44	0.79	27948.3	20172.3	48120.6	0.48	0.79
(0.4,0.2)	29527.1	20062.6	49589.7	0.49	0.79	27684.0	20194.0	47878.0	0.38	0.80

4.4.5 Storage vs. No Storage

Here we compare the experimental results to show the effects of ES installation. From previous experiments, we observe that the PSAA model produces higher-quality solutions than the SAA model. Thus, all subsequent experiments are conducted using the PSAA model.

In the experiments, the maximum running time $\vartheta = 12$ hours for the IEEE 33-Bus system and $\vartheta = 16$ hours for the IEEE 123-Bus system, and the planning horizon $T = 3$ weeks. Note that our proposed TCC model (4.7) is general enough to consider a longer-term setting because one can always set T to be years or seasons. Here our experiments consider a representative snapshot of the long-term future by setting T to be a relatively small number. Correspondingly, the cost parameters in the first stage of model (4.7), including the setup costs and the size-based investment/maintenance costs of the RDG and ES units, have also been levelized over the specific T time periods (weeks).

To match the planning horizon, we integrate every 3 of the 4000 weekly samples into a 3-week-long sample (without repetition) to obtain 1333 new samples. For computational efficiency, the size of the training data is set to $\Pi_1 = \Pi_2 = 50$ for the IEEE 33-Bus system and $\Pi_1 = \Pi_2 = 30$ for the IEEE 123-Bus system. As the training set is relatively small compared with the number of random variables, we use the k-means clustering algorithm to improve the reliability of the training samples. In particular, we randomly choose 200 of the 1333 samples and divide them into 50 groups by the k-means algorithm. We then use the centers of the 50 groups as our training samples. The remaining 1133 samples are used for testing.

We modify the active-power upper bounds of the DDG units (i.e., $\bar{p}_n^t$) and the standard capacities in $\mathcal{X}$ of the RDG units to match their designed load share in different cases. In particular, let ϱ_1 and ϱ_2 be two nonnegative parameters, and let SUM be the expected total active load (estimated from the real data and invariant to t). For each of the two DDG units, the active-power upper bound $\bar{p}_n^t$ is adjusted to $\varrho_1 \times SUM/2$ for all

$t \in [T]$. The maximum standard capacity of the RDG units $\bar{x}_4$ is adjusted from 7 MW to $\varrho_2 \times SUM/4$. The other three standard capacities in $\mathcal{X}$ are adjusted proportionally. For example, the minimum standard capacity $\bar{x}_1$ is adjusted from 4 MW to $\varrho_2 \times SUM/4 \times 4/7$. We conduct two sets of experiments with $\eta = 0.1$ and $\eta = 0.2$, respectively. For each set of experiments, we set (ϱ_1, ϱ_2) to take six different pairs of values, i.e., $(0.5, 0.5)$, $(0.4, 0.6)$, $(0.6, 0.4)$, $(0.3, 0.3)$, $(0.2, 0.4)$, and $(0.4, 0.2)$. Note that (ϱ_1, ϱ_2) does not represent an interval. When $(\varrho_1, \varrho_2) = (0.4, 0.6)$, it means that we set $\varrho_1 = 0.4$ and $\varrho_2 = 0.6$. The results are shown in Tables 4.5 – 4.8.

When $\varrho_1 + \varrho_2 = 1.0$, the solutions with ES units lead to higher second-stage costs but lower total costs than those without ES units. This indicates that ES installation is beneficial overall, despite leading to higher operational costs. Regarding computational efficiency, the optimality gaps for the solutions with ES are no smaller than those without ES, but the difference is negligible. Thus, considering ES installation increases the problem complexity, but not significantly.

When $\varrho_1 + \varrho_2 = 0.6$ (i.e., the DDG and RDG units may be insufficient to satisfy the load), ES units are more crucial. Without ES units, we cannot find a feasible solution within the time limit in any case. This indicates that the problems without ES units are likely to be infeasible. However, with ES units, feasible solutions are found within the time limit, and the actual feasible probabilities of the solutions are very close to the desired probabilities. This is because more active power can be purchased or generated in advance when there are ES units, and thus fulfill the load when the demand is high.

As an illustrative example, we consider the instance with $(\varrho_1, \varrho_2) = (0.6, 0.4)$ and $\eta = 0.2$ in the IEEE 33-Bus system. Without ES units, the solution sites (sizes) candidate RDG units at Buses 4 (7 MW), 9 (6 MW), and 28 (7 MW). With ES units, the result shows that candidate RDG units should be installed at Buses 3 (5 MW), 12 (3 MW), and 29 (7 MW), and that ES units should be installed at Buses 2 (2.3 MW), 12 (1.8 MW), and 27 (2.8 MW). The latter solution agrees with our intuition that storage units placed close to high-load buses play an important role in balancing the supply and demand in the power grid. We also observe that ES installation reduces the total required capacities of RDG units. This reduction lowers the first-stage costs so much that the total costs are reduced, even though the operating costs increase due to the operation of storage units.

Finally, all the above numerical results demonstrate that our proposed TCC model and PSAA approach can effectively deal with the RDG and ES planning problem under significant uncertainties. We note that, although we focus on such a planning problem in this chapter, the proposed model and approach can also be applied to other practical problems under uncertainty in the industry. For instance, we can apply the PSAA approach to solve the chance-constrained unit commitment problems in Wang et al. (2012) and chance-constrained optimal power flow problems in Roald and Andersson (2017), thereby reducing computational challenge. In general, many practical problems that consider two-stage decision-making under uncertainty may be formulated as a TCC model and solved by the PSAA approach. In addition, our proposed PSAA approach is a data-driven approach because (i) we use historical data to represent the possible scenarios of uncertain parameters and accordingly characterize chance constraints in our model; (ii) we use historical data to estimate the cumulative distribution function of a single random parameter ξ_1' by a non-parametric estimation technique; and (iii) once we obtain the first-stage solution of our model, we use real data to test the effectiveness of the obtained solution, simulating real practices. Such an approach can be applied to a wider range of practical problems.

4.5 Conclusions

Distribution grid operators face great challenges in deciding the locations and capacities of RDG and ES units due to significant uncertainties and complexities of distribution systems (e.g., ACOPF). To support such a decision-making problem, we develop a novel

TCC model to ensure system reliability, minimize costs, and improve renewable energy penetration. One key feature of our model is that the chance constraint ensures that all the operational constraints are satisfied simultaneously with a high probability, leading to system reliability. We use two sampling techniques to reformulate our developed model, leading to the standard SAA formulation and our proposed PSAA formulation. The novelty of the PSAA formulation is that it introduces only continuous variables corresponding to the samples (as compared to integer variables in the SAA formulation) and uses historical data to improve its performance. Our extensive experiments show that the PSAA formulation performs better than the SAA formulation. The PSAA provides better locations and capacities of the RDG and ES units in a shorter time with a lower total cost and achieves a better desired probability of ensuring system feasibility than the SAA. The PSAA also reduces the optimality gap by more than 50% as compared to the SAA. We finally demonstrate the significance of ES units in reducing total costs and improving the power system balance.

This research can be extended in various directions. First, as our proposed TCC model and PSAA approach is general enough, it would be interesting to apply the TCC model and PSAA approach to solve other practical problems in power system planning and operations. Second, the PSAA approach always finds a better solution in a shorter time than the SAA approach in our numerical experiments, but we do not have a theoretical proof for such results. A theoretical study would be appealing. Third, although we consider a radial distribution network in this chapter, there can be other types of distribution networks, e.g., meshed grids (Farivar and Low 2013) and multiphase grids (Gan and Low 2014). One can apply various approximations (e.g., semidefinite programming (Low 2014)) to formulate the corresponding ACOPF constraints. Fourth, although we adopt Bender's decomposition algorithms to improve computational efficiency, more advanced algorithms can be developed. We leave them to future research.

Chapter 5

Conclusions

In this thesis, we present three studies on computationally efficient approaches for modern optimization problems.

The first study considers the moment-based DRO problem. Although most moment-based DRO problems can be reformulated as SDP problems that can be solved in polynomial time, solving high-dimensional SDPs is still time-consuming. Unlike existing approximation approaches that first reduce the dimensionality of random parameters and then solve the approximated SDPs, we propose an ODR approach by integrating the dimensionality reduction of random parameters with the subsequent optimization problems. Such integration enables two outer and one inner approximations of the original problem, all of which are low-dimensional SDPs that can be solved efficiently, providing two lower bounds and one upper bound correspondingly. More importantly, these approximations can theoretically achieve the optimal value of the original high-dimensional SDPs. As these approximations are nonconvex SDPs, we develop modified ADMM algorithms to solve them efficiently. We demonstrate the effectiveness of our proposed ODR approach and algorithm in solving multiproduct newsvendor and production-transportation problems. Numerical results show significant advantages of our approach regarding computational time and solution quality over the three best possible benchmark approaches.

Our approach can obtain an optimal or near-optimal (mostly within 0.1%) solution and reduce the computational time by up to three orders of magnitude. This chapter can be further extended in various directions. First, we consider a piece-wise linear cost function in the original problem. Thus, it would be attractive to consider a more general objective function. Second, it would be interesting to apply our approach to more application problems to generate practical insights. Third, it is also of great interest to integrate the idea of dimensionality reduction into the Wasserstein DRO or two-stage stochastic programming. Fourth, our ODR approach can be potentially generalized to solve general SDPs with certain structures. Thus, it would be appealing to exploit the structures of SDP constraints and apply the ODR approach to solve more general SDPs. Fifth, it would be an interesting extension to consider cases with $K \geq m$. Note that for these cases, our proposed ODR approach can still be used to improve the traditional PCA approximation by choosing an appropriate m_1 such that the solution quality is good and the computational time is short, and also provide theoretical guarantees that bound the gap between the proposed upper bound and the original optimal value, as well as the gap between the original optimal value and the proposed lower bound. Nevertheless, we may not provide clear guidance on choosing m_1 so that our proposed approximations achieve the optimal value of the original problem. We leave the above extensions for future research.

The second study explores the non-convex QCPs, which are generally NP-hard and challenging problems. We propose two novel MILP approximations for a non-convex QCP. Our method begins by utilizing an eigenvalue-based decomposition to express the non-convex quadratic function as the difference of two convex functions. We then introduce an additional variable to partition each non-convex constraint into an SOC constraint and the complement of an SOC constraint. We employ two polyhedral approximation approaches to approximate the SOC constraint. The complement of an SOC constraint is approximated using a combination of linear and complementarity constraints. As a result, we approximate the non-convex QCP with two LPCCs. More importantly, we prove that the optimal values of the LPCCs asymptotically converge to that of the original non-convex QCP. By proving the boundedness of the LPCCs, we further reformulate the LPCCs as MILPs. We demonstrate the effectiveness of our approaches via numerical experiments by applying our proposed approximations to randomly generated instances and two application problems: the joint decision and estimation problem and the two-trust-region subproblem. The numerical results show significant advantages of our approaches in terms of solution quality and computational time compared to existing benchmark approaches. This study can be extended in various directions. First, there are a lot of auxiliary variables in our proposed MILP approximations. It would be interesting to study how to reduce the number of these auxiliary variables. Second, Proposition 9 provides a relaxation for the original quadratic problem. It is interesting to quantify the difference between the relaxed problem and the original problem in terms of the optimal value and optimal solutions. Third, as our proposed MILP approximations are general enough, it would be interesting to apply them to solve other practical problems in various fields. We leave them to future research.

The third study considers the planning of distributed energy resources in distribution systems. This problem has been challenged by the significant uncertainties and complexities of distribution systems. To ensure system reliability, one often employs chance-constrained programs to seek a highly likely feasible solution while minimizing certain costs. The traditional SAA is commonly used to represent uncertainties and reformulate a chance-constrained program into a deterministic optimization problem. However, the SAA introduces additional binary variables to indicate whether a scenario sample is satisfied and thus brings great computational complexity to the already challenging distributed energy resource planning problems. In this study, we introduce a new paradigm, i.e., PSAA using real data, to improve computational tractability. The innovation is that we sample only a part of the random parameters and introduce only continuous variables corresponding to the samples in the reformulation, which is a mixed-integer convex quadratic program. Our extensive experiments on the IEEE 33-Bus and 123-Bus systems show that the PSAA approach performs better than the SAA because the former provides better solutions in a shorter time in in-sample tests and provides better guaranteed probability for system reliability in out-of-sample tests. This research can be extended in various directions. First, as our proposed TCC model and PSAA approach is general enough, it would be interesting to apply the TCC model and PSAA approach to solve other practical problems in power system planning and operations. Second, the PSAA approach always finds a better solution in a shorter time than the SAA approach in our numerical experiments, but we do not have a theoretical proof for such results. A theoretical study would be appealing. Third, although we consider a radial distribution network in this chapter, there can be other types of distribution networks, e.g., meshed grids (Farivar and Low 2013) and multiphase grids (Gan and Low 2014). One can apply various approximations (e.g., semidefinite programming (Low 2014)) to formulate the corresponding ACOPF constraints. Fourth, although we adopt Bender's decomposition algorithms to improve computational efficiency, more advanced algorithms can be developed. We leave them to future research.

Appendix A

Supplements for Chapter 2

A.1 Table of Notations

Table A.1: Summary of Notations

Notation	Description
Random Variables:	
$\boldsymbol{\xi}$	The random vector $\boldsymbol{\xi} \in \mathbb{R}^m$
$\boldsymbol{\xi}_{\mathrm{I}}$	The random vector $\boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m$ obtained by the linearly transformation of $\boldsymbol{\xi}$
$\boldsymbol{\xi}_{\mathrm{r}}$	The random vector $\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1}$ obtained by reducing the dimension of $\boldsymbol{\xi}_{\mathrm{I}}$
Distributions:	
$\mathbb{P}$	The probability distribution of the random vector $\boldsymbol{\xi}$
$\mathbb{P}_{\mathrm{I}}$	The probability distribution of the random vector $\boldsymbol{\xi}_{\mathrm{I}}$
$\mathbb{P}_{\mathrm{r}}$	The probability distribution of the random vector $\boldsymbol{\xi}_{\mathrm{r}}$
Decision	

Variables:	
$\mathbf{x}$	The decision variable $\mathbf{x} \in \mathbb{R}^n$
s, $\boldsymbol{\lambda}_k$, $\mathbf{q}$, $\mathbf{Q}$	Decision variables in original SDP problem
$\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{\lambda}} := \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\}$
$\mathbf{q}_\mathrm{r}$, $\mathbf{Q}_\mathrm{r}$	Decision variables in PCA approximation
$\mathbf{B}$	The decision variable used in the optimized dimensionality reduction
$t_k, \mathbf{p}_k, \mathbf{P}_k, \mathbf{Z}$	Decision variables used in the lower bound
$\mathbf{Q}'_\mathrm{r}, \hat{\mathbf{u}}', \hat{\mathbf{u}}'', \mathbf{B}_1, \mathbf{B}_2$	Decision variables used in the revisited lower bound
Parameters and Sets:	
$\mathcal{X}$	The feasible set of decision variable $\mathbf{x}$
$\mathcal{D}_{\mathrm{M0}}$	The distributional ambiguity set constructed by statistical information
$\mathcal{D}_\mathrm{M}$	The distributional ambiguity set corresponding to $\boldsymbol{\xi}_\mathrm{I}$
$\mathcal{S}$	The support of $\boldsymbol{\xi}$
γ_1	A scalar $\gamma_1 \geq 0$
γ_2	A scalar $\gamma_2 \geq 1$
$\boldsymbol{\mu}$	The estimated mean of $\boldsymbol{\xi}$
$\boldsymbol{\Sigma}$	The estimated covariance matrix of $\boldsymbol{\xi}$
$\mathbf{U}$, $\boldsymbol{\Lambda}$	Two matrices produced by the eigenvalue decomposition on the covariance matrix $\boldsymbol{\Sigma}$
$\mathbf{A}, \mathbf{b}$	$\mathcal{S} := \{\boldsymbol{\xi} \mid \mathbf{A}\boldsymbol{\xi} \leq \mathbf{b}\}$
$\mathcal{S}_\mathrm{I}$	The support of $\boldsymbol{\xi}_\mathrm{I}$
$\mathcal{S}_\mathrm{r}$	The support of $\boldsymbol{\xi}_\mathrm{r}$
$\mathcal{D}_\mathrm{L}$	The distributional ambiguity set corresponding to $\boldsymbol{\xi}_\mathrm{r}$
$\mathcal{B}_{m_1}$	The feasible set of decision variable $\mathbf{B} \in \mathbb{R}^{m \times m_1}$
$\mathcal{D}_\mathrm{U}$	The distributional ambiguity set by relaxing the second-moment constraint in $\mathcal{D}_\mathrm{M}$

Optimal Value

Functions:

$\Theta_{\mathrm{M}}(m)$	The optimal value of the original problem
$\Theta_{\mathrm{M}}(m_1)$	The optimal value of the PCA approximation
$\Theta_{\mathrm{L}}(m_1)$	The optimal value of the first outer approximation
$\underline{\Theta}(m_1, \mathbf{B})$	The optimal value of the subproblem of the first outer approximation
$\Theta_{\mathrm{U}}(m_1)$	The optimal value of the inner approximation
$\overline{\Theta}(m_1, \mathbf{B})$	The optimal value of the subproblem of the inner approximation
$\Theta_{\mathrm{L2}}(m_1)$	The optimal value of the second outer approximation

A.2 Supplement to Section 2.2

A.2.1 Reformulations in Example 1

The distributionally robust counterpart of the CVaR problem (2.8) can be formulated as

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{D}} \min_{t\in\mathbb{R}} t + \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}\left[g\left(\mathbf{x},\boldsymbol{\xi}\right) - t\right]^{+} \\ =& \min_{\mathbf{x}\in\mathcal{X},\ t\in\mathbb{R}} \max_{\mathbb{P}\in\mathcal{D}} t + \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}\left[g\left(\mathbf{x},\boldsymbol{\xi}\right) - t\right]^{+}, \end{aligned} \tag{A.1}$$

where the equality holds by the Sion's minimax theorem (Sion 1958) because $t + (1/\alpha)\mathbb{E}_{\mathbb{P}}[g(\mathbf{x},\boldsymbol{\xi}) - t]^{+}$ is convex in t, concave (specifically, linear) in $\mathbb{P}$, and $\mathcal{D}$ is compact. By Proposition 1, Problem (A.1) has the same optimal value with the following SDP formulation:

$$\min_{\substack{\mathbf{x},s,t,\boldsymbol{\lambda}_1,\\ \boldsymbol{\lambda}_2,\mathbf{q},\mathbf{Q}}} \quad s + \mathbf{I}_m \bullet \mathbf{Q} \tag{A.2a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - t - \boldsymbol{\lambda}_1^\top \mathbf{b} + \boldsymbol{\lambda}_1^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{A.2b}$$

$$\begin{bmatrix} s-\left(1-\frac{1}{\alpha}\right)t-\boldsymbol{\lambda}_2^\top \mathbf{b}-\left(\frac{1}{\alpha}\mathbf{x}\right)^\top \boldsymbol{\mu}+\boldsymbol{\lambda}_2^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2-\frac{1}{\alpha}\mathbf{x}\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2-\frac{1}{\alpha}\mathbf{x}\right)\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{A.2c}$$

$$\mathbf{x}\in\mathcal{X},\ t\in\mathbb{R},\ \boldsymbol{\lambda}_1\in\mathbb{R}_+^l,\ \boldsymbol{\lambda}_2\in\mathbb{R}_+^l. \tag{A.2d}$$

A.3 Supplement to Section 2.3

A.3.1 Proof of Lemma 1

First, we have

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0 \iff \mathbf{I}_m-\mathbf{B}\mathbf{I}_{m_1}^{-1}\mathbf{B}^\top \succeq 0 \iff \mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m,$$

where the first equivalence is by Schur complement and the second is because $\mathbf{I}_{m_1}^{-1}=\mathbf{I}_{m_1}$.

Second, we have

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{I}_{m_1} \end{bmatrix} \succeq 0 \iff \mathbf{I}_{m_1}-\mathbf{B}^\top\mathbf{I}_{m}^{-1}\mathbf{B} \succeq 0 \iff \mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1},$$

where the first equivalence is by Schur complement and the second is because $\mathbf{I}_{m}^{-1}=\mathbf{I}_{m}$. Thus, the lemma is proved. □

A.3.2 Proof of Lemma 2

(i) Suppose $\mathbf{X}\succeq\mathbf{Y}$. For any $\mathbf{a}\in\mathbb{R}^n$, we have $\mathbf{V}\mathbf{a}\in\mathbb{R}^m$. It follows that

$$\begin{aligned} \mathbf{X}\succeq\mathbf{Y} &\Longrightarrow (\mathbf{V}\mathbf{a})^\top(\mathbf{X}-\mathbf{Y})(\mathbf{V}\mathbf{a})\geq 0,\ \forall \mathbf{a}\in\mathbb{R}^n \\ &\iff \mathbf{a}^\top\left(\mathbf{V}^\top(\mathbf{X}-\mathbf{Y})\mathbf{V}\right)\mathbf{a}\geq 0,\ \forall \mathbf{a}\in\mathbb{R}^n \end{aligned}$$

$$\iff \mathbf{V}^\top(\mathbf{X}-\mathbf{Y})\mathbf{V} \succeq 0 \iff \mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}.$$

(ii) First, for any $\mathbf{V} \in \mathbb{R}^{m\times m}$, we have

$$\mathbf{X} \succeq \mathbf{Y} \Longrightarrow \mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$$

by (i). Second, suppose $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$. Note that $\mathbf{V}^{-1} \in \mathbb{R}^{m\times m}$. According to (i), the matrix $\mathbf{V}^\top\mathbf{X}\mathbf{V} - \mathbf{V}^\top\mathbf{Y}\mathbf{V}$ remains as PSD if it multiplies $(\mathbf{V}^{-1})^\top$ before it and $\mathbf{V}^{-1}$ after it, i.e.,

$$\left(\mathbf{V}^{-1}\right)^\top \mathbf{V}^\top\mathbf{X}\mathbf{V}\mathbf{V}^{-1} \succeq \left(\mathbf{V}^{-1}\right)^\top \mathbf{V}^\top\mathbf{Y}\mathbf{V}\mathbf{V}^{-1}.$$

It follows that $\mathbf{X} \succeq \mathbf{Y}$ because $(\mathbf{V}^{-1})^\top\mathbf{V}^\top = \mathbf{I}_m$ and $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}_m$. Thus, $\mathbf{X} \succeq \mathbf{Y}$ is equivalent to $\mathbf{V}^\top\mathbf{X}\mathbf{V} \succeq \mathbf{V}^\top\mathbf{Y}\mathbf{V}$ if $\mathbf{V} \in \mathbb{R}^{m\times m}$ is invertible. □

A.3.3 Proof of Theorem 1

(i) Given any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we define $\boldsymbol{\zeta} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu}$ and use $\mathcal{S}_\zeta$ and $\mathcal{D}_\zeta$ to denote its support and ambiguity set, respectively. As $\mathcal{S}_{\mathrm{r}} = \{\boldsymbol{\xi}_{\mathrm{r}} \in \mathbb{R}^{m_1} \mid \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu} \in \mathcal{S}\}$ and $\mathcal{S}_\zeta = \{\boldsymbol{\zeta} \in \mathbb{R}^m \mid \boldsymbol{\zeta} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}} + \boldsymbol{\mu},\ \boldsymbol{\xi}_{\mathrm{r}} \in \mathcal{S}_{\mathrm{r}}\}$, we can deduce $\mathcal{S}_\zeta \subseteq \mathcal{S}$. We also have

$$\begin{aligned}
&\left(\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbb{E}_{\mathbb{P}_\zeta}[\boldsymbol{\zeta}] - \boldsymbol{\mu}\right) = \left(\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right]\right)^\top \boldsymbol{\Sigma}^{-1}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_{\mathrm{r}}\right] \\
&= \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^\top\right]\mathbf{B}^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^\top\right]\mathbf{B}^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^\top\right]\mathbf{B}^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}] \\
&= \mathbb{E}_{\mathbb{P}_{\mathrm{r}}}\left[\boldsymbol{\xi}_{\mathrm{r}}^\top\right]\mathbf{B}^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)\mathbf{B}\mathbb{E}_{\mathbb{P}_{\mathrm{r}}}[\boldsymbol{\xi}_{\mathrm{r}}]
\end{aligned}$$

$$= \mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[\boldsymbol{\xi}_\mathrm{r}^\top\right]\mathbf{B}^\top\mathbf{B}\mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[\boldsymbol{\xi}_\mathrm{r}\right] = \mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[\boldsymbol{\xi}_\mathrm{r}^\top\right]\mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[\boldsymbol{\xi}_\mathrm{r}\right] \leq \gamma_1, \tag{A.3}$$

where the inequality holds because of (2.6b). Meanwhile, we have

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_\zeta}\left[(\boldsymbol{\zeta}-\boldsymbol{\mu})(\boldsymbol{\zeta}-\boldsymbol{\mu})^\top\right] = \mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_\mathrm{r}\boldsymbol{\xi}_\mathrm{r}^\top\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top\right]\\ \preceq&\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\gamma_2\mathbf{I}_{m_1}\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top = \gamma_2\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\mathbf{B}^\top\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^\top \preceq \gamma_2\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \gamma_2\boldsymbol{\Sigma},\end{aligned} \tag{A.4}$$

where the first inequality holds because of (2.6b) and the second inequality holds because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, leading to $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$, which is further equivalent to $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$ by Lemma 1. By $\mathcal{S}_\zeta \subseteq \mathcal{S}$, (A.3), and (A.4), it follows that $\mathcal{D}_\zeta$ lies in $\mathcal{D}_{\mathrm{M0}}$, i.e., $\mathcal{D}_\zeta \subseteq \mathcal{D}_{\mathrm{M0}}$.

Therefore, given any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, we have

$$\max_{\mathbb{P}_\mathrm{r}\in\mathcal{D}_\mathrm{L}} \mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_\mathrm{r}+\boldsymbol{\mu}\right)\right] = \max_{\mathbb{P}_\zeta\in\mathcal{D}_\zeta} \mathbb{E}_{\mathbb{P}_\zeta}\left[f\left(\mathbf{x},\boldsymbol{\zeta}\right)\right] \leq \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}}\left[f\left(\mathbf{x},\boldsymbol{\xi}\right)\right],$$

where the equality holds by change of variables and the inequality holds because $\mathcal{D}_\zeta \subseteq \mathcal{D}_{\mathrm{M0}}$. It follows that

$$\max_{\mathbf{B}\in\mathcal{B}_{m_1}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_\mathrm{r}\in\mathcal{D}_\mathrm{L}} \mathbb{E}_{\mathbb{P}_\mathrm{r}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\boldsymbol{\xi}_\mathrm{r}+\boldsymbol{\mu}\right)\right] \leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{D}_{\mathrm{M0}}} \mathbb{E}_{\mathbb{P}}\left[f\left(\mathbf{x},\boldsymbol{\xi}\right)\right],$$

which demonstrates that the optimal value of Problem (2.10) is a lower bound for that of Problem (2.3) (i.e., Problem (2.2)).

(ii) For any $m_1 < m_2 \leq m$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, and $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $\mathbf{B}_1^\top\mathbf{B}_1 = \mathbf{I}_{m_1}$ and $[\mathbf{B}_1,\mathbf{C}]^\top[\mathbf{B}_1,\mathbf{C}] = \mathbf{I}_{m_2}$, we have $\mathbf{B}_2 = [\mathbf{B}_1,\mathbf{C}] \in \mathbb{R}^{m\times m_2}$. Meanwhile, we have $\mathcal{B}_{m_2} = \{\mathbf{B} \in \mathbb{R}^{m\times m_2} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_2}\}$ and define $\boldsymbol{\zeta}_i = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_i\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu} \in \mathbb{R}^m$ for any $i \in [2]$, where $\boldsymbol{\xi}_{\mathrm{r}_i} \in \mathbb{R}^{m_i}$. Clearly, $\mathbf{B}_2 \in \mathcal{B}_{m_2}$ because $\mathbf{B}_2^\top\mathbf{B}_2 = \mathbf{I}_{m_2}$. We further define the ambiguity set of $\boldsymbol{\zeta}_i$ as

$$\mathcal{D}_{\zeta_i} = \left\{\mathbb{P}_{\zeta_i} \;\middle|\; \boldsymbol{\zeta}_i \sim \mathbb{P}_{\zeta_i},\ \boldsymbol{\zeta}_i = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_i\boldsymbol{\xi}_{\mathrm{r}_i} + \boldsymbol{\mu},\ \boldsymbol{\xi}_{\mathrm{r}_i} \sim \mathbb{P}_{\mathrm{r}_i} \in \mathcal{D}_{\mathrm{r}_i}\right\},\ \forall i \in [2], \tag{A.5}$$

where $\mathcal{D}_{\mathrm{r}_i}$ represents the ambiguity set of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i \in [2]$. Given $\boldsymbol{\zeta}_1 \sim \mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_1}$, there exists a $\boldsymbol{\xi}_{\mathrm{r}_1} \sim \mathbb{P}_{\mathrm{r}_1} \in \mathcal{D}_{\mathrm{r}_1}$ such that $\boldsymbol{\zeta}_1 = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, where $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} = (\boldsymbol{\xi}_{\mathrm{r}_1}^\top, \mathbf{0}_{m_2-m_1}^\top)^\top \in \mathbb{R}^{m_2}$.

By using $\mathcal{S}_{\mathrm{r}_i}$ (see definition in (2.11)) to denote the support of $\boldsymbol{\xi}_{\mathrm{r}_i}$ for any $i \in [2]$, we have

$$\mathbb{P}\left\{\boldsymbol{\xi}_{\mathrm{r}_1} \in \mathcal{S}_{\mathrm{r}_1}\right\} = \mathbb{P}\left\{\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu} \in \mathcal{S}\right\} = \mathbb{P}\left\{\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu} \in \mathcal{S}\right\} = 1,$$

where the second equality holds because $\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$. It follows that $\mathbb{P}\{\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} \in \mathcal{S}_{\mathrm{r}_2}\} = 1$ by the definition of $\mathcal{S}_{\mathrm{r}_2}$. In addition, we have $\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top]\mathbb{E}[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}] = \mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}^\top]\mathbb{E}[\boldsymbol{\xi}_{\mathrm{r}_1}] \le \gamma_1$ and

$$\mathbb{E}\left[\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}^\top\right] = \begin{bmatrix} \mathbb{E}\left[\boldsymbol{\xi}_{\mathrm{r}_1}\boldsymbol{\xi}_{\mathrm{r}_1}^\top\right] & \mathbf{0}_{m_1\times(m_2-m_1)} \\ \mathbf{0}_{(m_2-m_1)\times m_1} & \mathbf{0}_{(m_2-m_1)\times(m_2-m_1)} \end{bmatrix} \preceq \gamma_2\mathbf{I}_{m_2}.$$

Thus, the probability distribution of $\bar{\boldsymbol{\xi}}_{\mathrm{r}_2}$ belongs to $\mathcal{D}_{\mathrm{r}_2}$. Meanwhile, by the definition of $\mathcal{D}_{\zeta_i}$ for any $i \in [2]$ in (A.5) and $\boldsymbol{\zeta}_1 = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\bar{\boldsymbol{\xi}}_{\mathrm{r}_2} + \boldsymbol{\mu}$, we have $\mathbb{P}_{\zeta_1} \in \mathcal{D}_{\zeta_2}$ and further $\mathcal{D}_{\zeta_1} \subseteq \mathcal{D}_{\zeta_2}$. Therefore, for any $\mathbf{x} \in \mathcal{X}$, $\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}$, and $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $\mathbf{B}_1^\top\mathbf{B}_1 = \mathbf{I}_{m_1}$ and $[\mathbf{B}_1, \mathbf{C}]^\top[\mathbf{B}_1, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\max_{\mathbb{P}_{\zeta_1}\in\mathcal{D}_{\zeta_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \boldsymbol{\zeta}_1\right)\right] \le \max_{\mathbb{P}_{\zeta_2}\in\mathcal{D}_{\zeta_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \boldsymbol{\zeta}_2\right)\right]. \tag{A.6}$$

Together with the definitions of $\boldsymbol{\zeta}_i$ ($\forall i \in [2]$) and $\mathbf{B}_2$, inequality (A.6) leads to

$$\max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \le \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right].$$

Considering an optimal solution $\mathbf{B}_1^* \in \mathbb{R}^{m\times m_1}$ of Problem (2.10), for any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $[\mathbf{B}_1^*, \mathbf{C}]^\top[\mathbf{B}_1^*, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \le \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right].$$

For any $\mathbf{C} \in \mathbb{R}^{m\times(m_2-m_1)}$ such that $[\mathbf{B}_1^*, \mathbf{C}]^\top [\mathbf{B}_1^*, \mathbf{C}] = \mathbf{I}_{m_2}$, we have

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right]. \tag{A.7}$$

It follows that

$$\begin{aligned}
&\max_{\mathbf{B}_1^\top\mathbf{B}_1=\mathbf{I}_{m_1}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \\
&= \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_1}\in\mathcal{D}_{\mathrm{r}_1}} \mathbb{E}_{\mathbb{P}_{\zeta_1}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1^*\boldsymbol{\xi}_{\mathrm{r}_1} + \boldsymbol{\mu}\right)\right] \\
&\leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}[\mathbf{B}_1^*, \mathbf{C}]\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right] \\
&\leq \max_{\mathbf{B}_2\in\mathcal{B}_{m_2}} \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}_{\mathrm{r}_2}\in\mathcal{D}_{\mathrm{r}_2}} \mathbb{E}_{\mathbb{P}_{\zeta_2}} \left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\boldsymbol{\xi}_{\mathrm{r}_2} + \boldsymbol{\mu}\right)\right],
\end{aligned}$$

where the first inequality holds by (A.7) and the second inequality holds because $[\mathbf{B}_1^*, \mathbf{C}] \in \mathcal{B}_{m_2}$. That is, the optimal value of Problem (2.10) is nondecreasing in m_1.

(iii) When $m_1 = m$, we have $\mathbf{B} \in \mathcal{B}_m \subseteq \mathbb{R}^{m\times m}$, i.e., $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_m$. First, we have $\Theta_{\mathrm{L}}(m) \leq \Theta_{\mathrm{M}}(m)$ by the conclusion (i). Second, when $\mathbf{B} = \mathbf{I}_m$, Problem (2.10) becomes Problem (2.3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (2.10), it follows that $\Theta_{\mathrm{L}}(m) \geq \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{L}}(m) = \Theta_{\mathrm{M}}(m)$. □

A.3.4 Theorem 9

Theorem 9. *Consider the optimal solution $(\mathbf{x}^*, s^*, \hat{\boldsymbol{\lambda}}^*, \mathbf{q}', \mathbf{Q}')$ of Problem (2.4), $S_k (\forall k \in [K])$, $\mathbf{V}$, $\boldsymbol{\delta}$, $\boldsymbol{\nu}_k (\forall k \in [K])$, and $\mathbf{Y}_{11}$ that are defined in Theorem 2. When $m_1 \geq K$, there exists a feasible solution $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$ in Problem (2.12) with $\mathbf{C} \in \mathbb{R}^{m\times(m_1-K)}$ and $[\mathbf{V}, \mathbf{C}]^\top[\mathbf{V}, \mathbf{C}] = \mathbf{I}_{m_1}$ and given this $\mathbf{B}^\dagger$, there exists a feasible solution $(\mathbf{x}^\dagger = \mathbf{x}^*, s^\dagger = s^*, \hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*, \mathbf{q}_{\mathrm{r}}^\dagger = (\boldsymbol{\delta}^\top, \mathbf{0}_{m_1-K}^\top)^\top, \mathbf{Q}_{\mathrm{r}}^\dagger = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{0}_{K\times(m_1-K)} \\ \mathbf{0}_{(m_1-K)\times K} & \mathbf{0}_{(m_1-K)\times(m_1-K)} \end{bmatrix})$ in Problem (2.13) such that the corresponding objective value equals the optimal value of Problem (2.4), $\Theta_{\mathrm{M}}(m)$.*

Proof. Proof. We construct a solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ of Problems (2.12) and (2.13) by setting $\mathbf{x}^\dagger = \mathbf{x}^*$, $s^\dagger = s^*$, $\hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*$, $\mathbf{q}_\mathrm{r}^\dagger = (\boldsymbol{\delta}^\top, \mathbf{0}_{m_1-K}^\top)^\top$, $\mathbf{Q}_\mathrm{r}^\dagger = \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{0}_{K\times(m_1-K)} \\ \mathbf{0}_{(m_1-K)\times K} & \mathbf{0}_{(m_1-K)\times(m_1-K)} \end{bmatrix}$, and $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$, where $\mathbf{C} \in \mathbb{R}^{m\times(m_1-K)}$ and $[\mathbf{V}, \mathbf{C}]^\top[\mathbf{V}, \mathbf{C}] = \mathbf{I}_{m_1}$. First, we show this constructed solution is feasible to Problems (2.12) and (2.13). Clearly, this solution satisfies constraints (2.13c). In addition, from Problem (2.4), as $\mathbf{q}' = \mathbf{V}\boldsymbol{\delta}$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, for any $k \in [K]$, we have

$$\begin{bmatrix} S_k & \frac{1}{2}\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right)^\top \\ \frac{1}{2}\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*))\right) & \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \end{bmatrix} \succeq 0,$$

which, by Schur complement, is equivalent to

$$S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\right) \succeq \frac{1}{4}\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top. \tag{A.8}$$

From (A.8), for any $k \in [K]$, we have the following inequality holds by Lemma 2:

$$\begin{aligned} &S_k\left([\mathbf{V}, \mathbf{C}]^\top\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top[\mathbf{V}, \mathbf{C}]\right) \\ &\succeq \frac{1}{4}[\mathbf{V}, \mathbf{C}]^\top\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top[\mathbf{V}, \mathbf{C}], \end{aligned}$$

which is equivalent to

$$S_k\mathbf{Q}_\mathrm{r}^\dagger \succeq \frac{1}{4}\left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{q}_\mathrm{r}^\dagger + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}^\dagger\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \tag{A.9}$$

by the construction of the solution $\mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger$ and $[\mathbf{V}, \mathbf{C}]^\top\mathbf{V} = [\mathbf{I}_K, \mathbf{0}_{K\times(m_1-K)}]^\top$. By Schur complement, (A.9) indicates that the constructed solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ also satisfies constraints (2.13b) and thus it is a feasible solution of Problems (2.12) and (2.13).

Second, we show the objective value of this feasible solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_\mathrm{r}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \mathbf{B}^\dagger)$ is equal to the optimal value of Problem (2.4). The objective value corresponding to this solution is

$$\begin{aligned}
s^\dagger + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^\dagger + \sqrt{\gamma_1}\left\|\mathbf{q}_{\mathrm{r}}^\dagger\right\|_2 &= s^* + \gamma_2 \mathbf{I}_K \bullet \mathbf{Y}_{11} + \sqrt{\gamma_1}\|\boldsymbol{\delta}\|_2 \\
&= s^* + \gamma_2 \mathbf{I}_K \bullet (\mathbf{Y}_{11}\mathbf{V}^\top\mathbf{V}) + \sqrt{\gamma_1}\|\boldsymbol{\delta}\|_2 \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet (\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) + \sqrt{\gamma_1}\|\boldsymbol{\delta}\|_2 \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\boldsymbol{\delta}\|_2 \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\boldsymbol{\delta}^\top\boldsymbol{\delta}} \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\boldsymbol{\delta}^\top\mathbf{V}^\top\mathbf{V}\boldsymbol{\delta}} \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\sqrt{\mathbf{q}'^\top\mathbf{q}'} \\
&= s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\|\mathbf{q}'\|_2 \\
&= \Theta_{\mathrm{M}}(m),
\end{aligned} \tag{A.10}$$

where the first equality holds by the construction of $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}_{\mathrm{r}}^\dagger, \mathbf{Q}_{\mathrm{r}}^\dagger, \mathbf{B}^\dagger)$, the second equality holds because $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K$, the third equality holds by the cyclic property of a matrix's trace, the fourth equality holds by the definition of $\mathbf{Q}'$ in Theorem 2, and the seventh equality holds because $\mathbf{q}' = \mathbf{V}\boldsymbol{\delta}$. □

Theorem 9 shows that, when $m_1 \geq K$, we can always find a feasible solution of Problems (2.12) and (2.13) such that the corresponding objective value is equal to the optimal value of the original Problem (2.4). More importantly, the SDP constraints in Problem (2.12) have smaller sizes (i.e., $m_1 + 1$) than those in Problem (2.4) (i.e., $m + 1$), potentially reducing computational challenges because K is usually small (e.g., $K = 2$ in the distributionally robust CVaR problem in Example 1).

It is important to note that, although the constructed feasible solution of Problem (2.12) has an objective value of $\Theta_{\mathrm{M}}(m)$ and Problem (2.12) serves as an outer approximation for Problem (2.4), this does not imply that the constructed feasible solution is optimal for Problem (2.12). The primary reason is that Problem (2.12) is a max-min problem. For any $m_1 \leq m$, we would like to have an optimal $\mathbf{B}^* \in \mathcal{B}_{m_1}$ such that the optimal value of the inner minimization problem given this $\mathbf{B}^*$ can be maximized. By Theorem 1, we have $\underline{\Theta}(m_1, \mathbf{B}^*) = \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}(m)$ for any $m_1 \leq m$.

According to Theorem 9, when $m_1 \geq K$, we can construct a feasible solution $\mathbf{B}^\dagger = [\mathbf{V}, \mathbf{C}]$ of the outer maximization problem. Note that $\mathbf{B}^\dagger$ is not necessarily optimal for Problem (2.12). Because the outer problem is a maximization problem, we have $\underline{\Theta}(m_1, \mathbf{B}^\dagger) \leq \underline{\Theta}(m_1, \mathbf{B}^*) = \Theta_{\mathrm{L}}(m_1) \leq \Theta_{\mathrm{M}}(m)$.

Given this feasible $\mathbf{B}^\dagger$, we can construct a feasible solution of the inner minimization problem (2.13) such that the corresponding objective value (here denoted by ϕ) is equal to $\Theta_{\mathrm{M}}(m)$. Note that this constructed feasible solution is not necessarily optimal for Problem (2.12) or Problem (2.13). Because the inner problem is a minimization problem, we have $\underline{\Theta}(m_1, \mathbf{B}^\dagger) \leq \phi = \Theta_{\mathrm{M}}(m)$.

Now, we construct a feasible solution $\mathbf{B}^\dagger$ for the outer maximization problem and, given this $\mathbf{B}^\dagger$, we also construct a feasible solution for the inner minimization problem. Although the corresponding objective value ϕ is equal to $\Theta_{\mathrm{M}}(m)$ and Problem (2.12) is an outer approximation for the original problem, we cannot claim that this constructed feasible solution is optimal due to the max-min natrue. One key reason is that the constructed solution of the inner minimization problem (2.13) may not be optimal. Thus, $\underline{\Theta}(m_1, \mathbf{B}^\dagger) < \phi = \Theta_{\mathrm{M}}(m)$ may hold.

We used to conjecture that this constructed feasible solution is an optimal solution of Problems (2.12) and (2.13) such that the optimal value of Problem (2.12) equals that of Problem (2.4) when $m_1 \geq K$. Most numerical experiments (see Section 2.7) show this conjecture may be correct, but we find a counter-example (see Example 2 in Appendix A.3.5). Example 2 illustrates that the optimal value of Problem (2.13) with $m_1 = K$ and $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (2.4), which means that the constructed feasible solution ($\mathbf{B} = \mathbf{V}$) is not optimal.

A.3.5 Counter-Example

Now we provide an example as follows to illustrate that the optimal value of Problem (2.13) with $m_1 = K$ and $\mathbf{B} = \mathbf{V}$ is strictly less than the optimal value of Problem (2.4), which means that the constructed feasible solution ($\mathbf{B} = \mathbf{V}$) is not optimal.

Example 2. *We consider an instance of Problem (2.4), where* $m = n = 4$, $\gamma_1 = 1$, $\gamma_2 = 2$, $\mathbf{A} = \mathbf{0}_{l\times m}$, $\mathbf{b} = \mathbf{0}_l$, $\boldsymbol{\mu} = \mathbf{1}_m$, $\boldsymbol{\Sigma} = \mathbf{I}_m$, $K = 3$, $y_k^0(\mathbf{x}) = 0$ $(\forall k \in [K])$, $y_k(\mathbf{x}) = \mathbf{W}_k\mathbf{x}$ $(\forall k \in [K])$ *with* $\mathbf{W}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\mathbf{W}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$, *and* $\mathbf{W}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, *and* $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^4 \mid x_1 = x_3 = x_4 = 1, x_2 \in \{-7, 1\}\}$, *then Problem (2.4) becomes*

$$\min_{\mathbf{x}\in\mathcal{X},s,\mathbf{q},\mathbf{Q}} \left\{ s + 2\mathbf{I}_m \bullet \mathbf{Q} + \|\mathbf{q}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top \mathbf{W}_k^\top \mathbf{1}_m & \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x})^\top \\ \frac{1}{2}(\mathbf{q} - \mathbf{W}_k\mathbf{x}) & \mathbf{Q} \end{bmatrix} \succeq 0, \ \forall k \in [K] \right\}. \tag{A.11}$$

Solving Problem (A.11) gives the optimal value 5.9882 *with* $\mathbf{x} = [1, 1, 1, 1]^\top$, $\mathbf{Q} = \begin{bmatrix} 0.0911 & -0.0558 & -0.0354 & -0.0558 \\ -0.0558 & 0.1115 & -0.0558 & 0.1115 \\ -0.0354 & -0.0558 & 0.0911 & -0.0558 \\ -0.0558 & 0.1115 & -0.0558 & 0.1115 \end{bmatrix}$, *and* $\operatorname{rank}(\mathbf{Q}) = 2$. *By Theorem 2, we can correspondingly obtain a feasible* $\mathbf{V} = \begin{bmatrix} 0.7071 & -0.5774 & -0.1543 \\ 0 & 0.5774 & -0.3086 \\ 0 & 0 & 0.9258 \\ 0.7071 & 0.5774 & 0.1543 \end{bmatrix}$. *Now given* $m_1 = K = 3$ *and* $\mathbf{B} = \mathbf{V}$, *Problem (2.13) becomes*

$$\min_{\mathbf{x}\in\mathcal{X},s,\mathbf{q}_{\mathrm{r}},\mathbf{Q}_{\mathrm{r}}} \left\{ s + 2\mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \|\mathbf{q}_{\mathrm{r}}\|_2 \;\middle|\; \begin{bmatrix} s - \mathbf{x}^\top \mathbf{W}_k^\top \mathbf{1}_m & \frac{1}{2}(\mathbf{q}_{\mathrm{r}} - \mathbf{V}^\top\mathbf{W}_k\mathbf{x})^\top \\ \frac{1}{2}(\mathbf{q}_{\mathrm{r}} - \mathbf{V}^\top\mathbf{W}_k\mathbf{x}) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \ \forall k \in [K] \right\}. \tag{A.12}$$

Solving Problem (A.12) gives the optimal value 5.1139 *with* $\mathbf{x} = [1, -7, 1, 1]^\top$. *That is, the optimal value of Problem (2.13) with* $\mathbf{B} = \mathbf{V}$ *is strictly less than the optimal value of*

Problem (2.4).

A.3.6 Proof of Proposition 2

We consider the Lagrangian dual of the inner minimization part (i.e., Problem (2.13)) of Problem (2.12) as follows:

$$\max_{\substack{\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0, \\ \forall k \in [K], \\ \mathbf{Z} \succeq 0}} \min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}} \geq 0, \\ \mathbf{q}_\mathrm{r}, \mathbf{Q}_\mathrm{r}}} \mathcal{L}\left(\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}_\mathrm{r}, \mathbf{Q}_\mathrm{r}; \mathbf{Z}, t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K]\right), \tag{A.13}$$

where the Lagrangian function

$$\begin{aligned}
&\mathcal{L}\left(\mathbf{x}, s, \hat{\boldsymbol{\lambda}}, \mathbf{q}_\mathrm{r}, \mathbf{Q}_\mathrm{r}; \mathbf{Z}, t_k, \mathbf{p}_k, \mathbf{P}_k, \forall k \in [K]\right) \\
=& s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1} \|\mathbf{q}_\mathrm{r}\|_2 - \mathbf{Z} \bullet \left(\sum_{i=1}^{n} (\boldsymbol{\Delta}_i x_i) + \boldsymbol{\Delta}_0\right) - \sum_{k=1}^{K} \begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \bullet \\
& \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top b - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A} \boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q}_\mathrm{r} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right)^\top \\ \frac{1}{2}\left(\mathbf{q}_\mathrm{r} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))\right) & \mathbf{Q}_\mathrm{r} \end{bmatrix} \\
=& \left(1 - \sum_{k=1}^{K} t_k\right) s - \sum_{k=1}^{K} \left(t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top\right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1} \|\mathbf{q}_\mathrm{r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_\mathrm{r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k\right) \bullet \mathbf{Q}_\mathrm{r} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \left(\mathbf{a}_{ij}\mathbf{x} + a_{ij}^0\right) \\
&+ \sum_{k=1}^{K} \left(t_k y_k^0(x) + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right) y_k(x)\right) \\
=& \left(1 - \sum_{k=1}^{K} t_k\right) s - \sum_{k=1}^{K} \left(t_k (\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \mathbf{A}^\top\right) \boldsymbol{\lambda}_k + \sqrt{\gamma_1} \|\mathbf{q}_\mathrm{r}\|_2 - \sum_{k=1}^{K} \mathbf{p}_k^\top \mathbf{q}_\mathrm{r} \\
&+ \left(\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k\right) \bullet \mathbf{Q}_\mathrm{r} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \mathbf{a}_{ij}\mathbf{x} - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} a_{ij}^0 \\
&+ \sum_{k=1}^{K} \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \mathbf{p}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right) \mathbf{W}_k\right) \mathbf{x}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=1}^{K}\left(t_k d_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)\\
=&\left(1-\sum_{k=1}^{K}t_k\right)s-\sum_{k=1}^{K}\left(t_k\left(\mathbf{A}\boldsymbol{\mu}-\mathbf{b}\right)^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\right)\boldsymbol{\lambda}_k+\sqrt{\gamma_1}\left\|\mathbf{q}_{\mathrm{r}}\right\|_2-\sum_{k=1}^{K}\mathbf{p}_k^\top\mathbf{q}_{\mathrm{r}}\\
&+\left(\gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^{K}\mathbf{P}_k\right)\bullet\mathbf{Q}_{\mathrm{r}}+\left(\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}\right)\mathbf{x}\\
&+\sum_{k=1}^{K}\left(t_k d_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}a_{ij}^0.
\end{aligned}
$$

To present the objective value of the inner minimization problem of (A.13) from going to negative infinity, we require

$$
1-\sum_{k=1}^{K}t_k=0,\ \sqrt{\gamma_1}-\left\|\sum_{k=1}^{K}\mathbf{p}_k\right\|_2\geq 0,\ \gamma_2\mathbf{I}_{m_1}-\sum_{k=1}^{K}\mathbf{P}_k=0, \tag{A.14a}
$$

$$
t_k\left(\mathbf{A}\boldsymbol{\mu}-\mathbf{b}\right)^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top\leq 0,\ \forall k\in[K], \tag{A.14b}
$$

$$
\sum_{k=1}^{K}\left(t_k\mathbf{w}_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{W}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}\mathbf{a}_{ij}=0, \tag{A.14c}
$$

$$
\begin{bmatrix} t_k & \mathbf{p}_k^\top\\ \mathbf{p}_k & \mathbf{P}_k\end{bmatrix}\succeq 0,\ \forall k\in[K],\ \mathbf{Z}\succeq 0. \tag{A.14d}
$$

Then, the dual problem of Problem (2.13) can be described as follows:

$$
\begin{aligned}
\max_{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\mathbf{Z}}\ &\sum_{k=1}^{K}\left(t_k d_k^0+\left(t_k\boldsymbol{\mu}^\top+\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)\mathbf{d}_k\right)-\sum_{i=1}^{\tau}\sum_{j=1}^{\tau}z_{ij}a_{ij}^0\\
\text{s.t.}\quad &(\text{A.14a})-(\text{A.14d}).
\end{aligned}
\tag{A.15}
$$

By integrating the outer maximization part of Problem (2.12) and Problem (A.15), we obtain the bilinear SDP problem (2.25). Now we would like to prove the strong duality between Problem (2.13) and Problem (A.15); that is, these two problems share the same optimal value, which further shows that Problem (2.12) has the same optimal value as Problem (2.25). To that end, we find an interior point of Problem (2.13).

Let $\mathbf{x}'$ be an interior point in $\mathcal{X}$, we can construct an interior point by setting $\hat{\boldsymbol{\lambda}}' = \{\mathbf{1}_l, \ldots, \mathbf{1}_l\}$, $s' = \sum_{k=1}^{K} \left| y_k^0(\mathbf{x}') + \mathbf{1}_l^\top \mathbf{b} + y_k(\mathbf{x}')^\top \boldsymbol{\mu} - \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu} \right| + 1$, $\mathbf{q}_{\mathrm{r}}' = 0$, and $\mathbf{Q}_{\mathrm{r}}' = \sum_{k=1}^{K} 1/(4(s' - y_k^0(\mathbf{x}') - \mathbf{1}_l^\top b - y_k(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu}))(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top (\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}) + \mathbf{I}_{m_1}$. Clearly, $(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top (\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))^\top (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}) \succeq 0$. Thus, $\mathbf{Q}_{\mathrm{r}}' \succ 0$. Now we only need to show that constraints (2.13b) hold in the positive-definite sense with respect to this constructed solution.

By the construction of $\mathbf{Q}_{\mathrm{r}}'$, for any $k \in [K]$, we have

$$
\begin{aligned}
&\mathbf{Q}_{\mathrm{r}}' - \frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}')\right)\right)^\top}{4\left(s' - y_k^0(\mathbf{x}') - \mathbf{1}_l^\top b - y_k(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu}\right)} \\
&= \sum_{\forall k' \in [K]: k' \neq k} \frac{\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_{k'}(\mathbf{x}')\right)\right)\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_{k'}(\mathbf{x}')\right)\right)^\top}{4\left(s' - y_{k'}^0(\mathbf{x}') - \mathbf{1}_l^\top b - y_{k'}(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu}\right)} + \mathbf{I}_{m_1} \succ 0,
\end{aligned}
\tag{A.16}
$$

where $s' - y_{k'}^0(\mathbf{x}') - \mathbf{1}_l^\top b - y_{k'}(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu} > 0$ by the construction of s'. By Schur complement, (A.16) is equivalent to

$$
\begin{bmatrix}
s' - y_k^0(\mathbf{x}') - \mathbf{1}_l^\top b - y_k(\mathbf{x}')^\top \boldsymbol{\mu} + \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}')\right)\right)^\top \\
\frac{1}{2}\left(\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}')\right)\right) & \mathbf{Q}_{\mathrm{r}}'
\end{bmatrix} \succ 0, \ \forall k \in [K].
$$

Thus, $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}_{\mathrm{r}}', \mathbf{Q}_{\mathrm{r}}')$ is an interior point of Problem (2.13) and the strong duality between Problem (2.13) and Problem (A.15) holds. □

A.4 Supplement to Section 2.4

A.4.1 Proof of Theorem 3

(i) For any $\boldsymbol{\xi}_{\mathrm{I}} \sim \mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}$, we have $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}] \preceq \gamma_2 \mathbf{I}_{m_1}$. Then, by Lemma 2, for any given $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, i.e., $\mathbf{B}^{\top}\mathbf{B} = \mathbf{I}_{m_1}$, we further have $\mathbf{B}^{\top}(\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}])\mathbf{B} \preceq \mathbf{B}^{\top}(\gamma_2\mathbf{I}_{m_1})\mathbf{B}$, i.e., $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[\mathbf{B}^{\top}\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}\mathbf{B}] \preceq \gamma_2\mathbf{B}^{\top}\mathbf{I}_{m_1}\mathbf{B} = \gamma_2\mathbf{I}_{m_1}$. It follows that $\mathcal{D}_{\mathrm{M}} \subseteq \mathcal{D}_{\mathrm{U}}$. Thus, given any $\mathbf{x} \in \mathcal{X}$ and $\mathbf{B} \in \mathcal{B}_{m_1}$, we have

$$\max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \geq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right].$$

It follows that

$$\min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \geq \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right],$$

which demonstrates that the optimal value of Problem (2.26) is an upper bound for that of Problem (2.3) (i.e., Problem (2.2)).

(ii) Consider any $m_1 < m_2 \leq m$. We have $\mathcal{B}_{m_2} := \{\mathbf{B} \in \mathbb{R}^{m \times m_2} \mid \mathbf{B}^{\top}\mathbf{B} = \mathbf{I}_{m_2}\}$ and consider an optimal solution $(\mathbf{B}^*, \mathbf{x}^*)$ of Problem (2.26), i.e., $\min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$.

Note that $(\mathbf{B}^*)^{\top}\mathbf{B}^* = \mathbf{I}_{m_1}$. We can then construct $\mathbf{B}' = [\mathbf{B}^*, \mathbf{C}] \in \mathbb{R}^{m \times m_2}$ such that $\mathbf{C} \in \mathbb{R}^{m \times (m_2 - m_1)}$ and $\mathbf{B}' \in \mathcal{B}_{m_2}$, i.e., $(\mathbf{B}')^{\top}\mathbf{B}' = \mathbf{I}_{m_2}$. With $\mathbf{B}'$, we use $\mathcal{D}'_{\mathrm{U}}$ to denote the corresponding ambiguity set defined in (2.28). By the second-moment constraint in $\mathcal{D}'_{\mathrm{U}}$, we have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[(\mathbf{B}')^{\top}\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}\mathbf{B}'\right] \\
=&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[[\mathbf{B}^*, \mathbf{C}]^{\top}\boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^{\top}[\mathbf{B}^*, \mathbf{C}]\right]
\end{aligned}$$

$$
\begin{aligned}
=&\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\begin{bmatrix} (\mathbf{B}^*)^\top \boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}^* & (\mathbf{B}^*)^\top \boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{C} \\ \mathbf{C}^\top \boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}^* & \mathbf{C}^\top \boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{C} \end{bmatrix} \\
\preceq&\gamma_2 \mathbf{I}_{m_2},
\end{aligned}
$$

which implies that $\mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[(\mathbf{B}^*)^\top \boldsymbol{\xi}_{\mathrm{I}}\boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}^*] \preceq \gamma_2 \mathbf{I}_{m_1}$. It follows that $\mathcal{D}'_{\mathrm{U}} \subseteq \mathcal{D}_{\mathrm{U}}$. Therefore, we have

$$
\max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \geq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right]. \tag{A.17}
$$

Because $\mathbf{B}' \in \mathcal{B}_{m_2}$ and $\mathbf{x}^* \in \mathcal{X}$, the constructed solution $(\mathbf{B}', \mathbf{x}^*)$ is feasible to the problem $\min_{\mathbf{B} \in \mathcal{B}_{m_2}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$. Then, we have

$$
\begin{aligned}
\Theta_{\mathrm{U}}(m_2) &= \min_{\mathbf{B} \in \mathcal{B}_{m_2}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&\leq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&\leq \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}^*, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&= \min_{\mathbf{B} \in \mathcal{B}_{m_1}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}\left[f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)\right] \\
&= \Theta_{\mathrm{U}}(m_1),
\end{aligned}
$$

where the first inequality holds because $(\mathbf{B}', \mathbf{x}^*)$ is a feasible solution of the problem $\min_{\mathbf{B} \in \mathcal{B}_{m_2}} \min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbb{P}_{\mathrm{I}} \in \mathcal{D}'_{\mathrm{U}}} \mathbb{E}_{\mathbb{P}_{\mathrm{I}}}[f(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu})]$, the second inequality holds by (A.17), and the second equality holds because $(\mathbf{B}^*, \mathbf{x}^*)$ is an optimal solution of Problem (2.26). That is, the optimal value of Problem (2.26) is nonincreasing in m_1.

(iii) When $m_1 = m$, we have $\mathbf{B} \in \mathcal{B}_m \subseteq \mathbb{R}^{m \times m}$, i.e., $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_m$. First, we have $\Theta_{\mathrm{U}}(m) \geq \Theta_{\mathrm{M}}(m)$ by the conclusion (i). Second, when $\mathbf{B} = \mathbf{I}_m$, Problem (2.26) becomes Problem (2.3). Because $\mathbf{B} = \mathbf{I}_m$ is a feasible solution of Problem (2.26), it follows that $\Theta_{\mathrm{U}}(m) \leq \Theta_{\mathrm{M}}(m)$. Therefore, we have $\Theta_{\mathrm{U}}(m) = \Theta_{\mathrm{M}}(m)$. $\square$

A.4.2 Proof of Proposition 3

First, by Theorem 3 in Cheramin et al. (2022), Problem (2.27) has the same optimal value as the following problem:

$$
\begin{aligned}
\min_{\mathbf{x},s,\mathbf{q},\mathbf{Q}_\mathrm{r}} \quad & s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\, \|\mathbf{q}\|_2 && \text{(A.18a)} \\
\text{s.t.} \quad & s \geq f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_\mathrm{I}^\top \mathbf{B}\mathbf{Q}_\mathrm{r}\mathbf{B}^\top \boldsymbol{\xi}_\mathrm{I} - \mathbf{q}^\top \boldsymbol{\xi}_\mathrm{I},\ \forall \boldsymbol{\xi}_\mathrm{I} \in \mathcal{S}_\mathrm{I}, && \text{(A.18b)} \\
& \mathbf{Q}_\mathrm{r} \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1 \times m_1},\ \mathbf{q} \in \mathbb{R}^m. && \text{(A.18c)}
\end{aligned}
$$

Next, we apply the strong duality theorem to constraints (A.18b). We define

$$
g_k(\boldsymbol{\xi}_\mathrm{I}) = s + \boldsymbol{\xi}_\mathrm{I}^\top \mathbf{B}\mathbf{Q}_\mathrm{r}\mathbf{B}^\top \boldsymbol{\xi}_\mathrm{I} + \mathbf{q}^\top \boldsymbol{\xi}_\mathrm{I} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right),\ \forall k \in [K].
$$

As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (A.18b) as

$$
g_k(\boldsymbol{\xi}_\mathrm{I}) \geq 0,\ \forall \boldsymbol{\xi}_\mathrm{I} \in \mathcal{S}_\mathrm{I},\ \forall k \in [K],
$$

which is equivalent to

$$
\min_{\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) \leq \mathbf{b},\ \boldsymbol{\xi}_\mathrm{I} \in \mathbb{R}^m} \quad g_k(\boldsymbol{\xi}_\mathrm{I}) \geq 0,\ \forall k \in [K]. \tag{A.19}
$$

For any $k \in [K]$, the Lagrangian dual problem of $\min_{\mathbf{A}(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}) \leq \mathbf{b},\ \boldsymbol{\xi}_\mathrm{I} \in \mathbb{R}^m} \ g_k(\boldsymbol{\xi}_\mathrm{I})$ is

$$
\max_{\boldsymbol{\lambda}_k \geq 0} \min_{\boldsymbol{\xi}_\mathrm{I} \in \mathbb{R}^m} \ g_k(\boldsymbol{\xi}_\mathrm{I}) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) - \mathbf{b}\right),
$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^l$. Because there exists an interior point for the primal problem, the strong duality holds. Thus, constraints (A.19) are equivalent to

$$
\max_{\boldsymbol{\lambda}_k \geq 0} \min_{\boldsymbol{\xi}_\mathrm{I}} \ g_k(\boldsymbol{\xi}_\mathrm{I}) + \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_\mathrm{I} + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0,\ \forall k \in [K],
$$

which are further equivalent to

$$\exists \boldsymbol{\lambda}_k \geq 0 : s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}\mathbf{Q}_{\mathrm{r}}\mathbf{B}^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right)$$
$$+ \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0, \ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m, \ \forall k \in [K]. \quad \text{(A.20)}$$

Note that $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$; that is, all the column vectors of $\mathbf{B}$ are orthogonal. We can then extend $\mathbf{B}$ to $[\mathbf{B}, \bar{\mathbf{B}}] \in \mathbb{R}^{m\times m}$ with $\bar{\mathbf{B}} \in \mathbb{R}^{m\times(m-m_1)}$ such that all the column vectors of $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$. Thus, we can always find $\boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1}$ such that

$$\boldsymbol{\xi}_{\mathrm{I}} = \mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2.$$

It follows that constraints (A.20) become

$$\exists \boldsymbol{\lambda}_k \geq 0 : s + \boldsymbol{\xi}_1^\top \mathbf{Q}_{\mathrm{r}}\boldsymbol{\xi}_1 + \mathbf{q}^\top \left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) + \boldsymbol{\mu}\right)$$
$$+ \boldsymbol{\lambda}_k^\top \left(\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}\boldsymbol{\xi}_1 + \bar{\mathbf{B}}\boldsymbol{\xi}_2\right) + \boldsymbol{\mu}\right) - \mathbf{b}\right) \geq 0, \ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1}, \ \forall k \in [K]. \quad \text{(A.21)}$$

We further define

$$\mathbf{Z}_k = \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{B}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right)^\top \\ \frac{1}{2}\left(\mathbf{B}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) & \mathbf{Q}_{\mathrm{r}} \end{bmatrix},$$
$$\forall k \in [K].$$

Thus, we have

$$\begin{aligned} &\text{(A.21)} \\ \iff\ & \exists \boldsymbol{\lambda}_k \geq 0 : \ \left(1, \boldsymbol{\xi}_1^\top\right)\mathbf{Z}_k\left(1, \boldsymbol{\xi}_1^\top\right)^\top + \boldsymbol{\xi}_2^\top\left(\bar{\mathbf{B}}^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) \geq 0, \end{aligned}$$

$$\forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{m-m_1},\ \forall k \in [K].$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0:\ \left(1, \boldsymbol{\xi}_1^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top\right)^\top \geq 0,\ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1},\ \forall k \in [K]; \tag{A.22}$$
$$\bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0,\ \forall k \in [K].$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0,\ \forall k \in [K].$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0:\ \mathbf{Z}_k \succeq 0,\ \bar{\mathbf{B}}^\top \left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) = 0,\ \forall k \in [K]. \tag{A.23}$$

$$\iff \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k \in \mathbb{R}^{m_1}:\ \mathbf{Z}_k \succeq 0,\ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}\mathbf{u}_k,\ \forall k \in [K]. \tag{A.24}$$

The first equivalence holds due to the definition of $\mathbf{Z}_k$. For the third equivalence, clearly $\Longleftarrow$ follows from the definition of a PSD matrix. To prove $\Longrightarrow$, we consider two possible cases for any $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}^\top \in \mathbb{R}^{m_1})^\top \in \mathbb{R}^{m_1+1}$: (i) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}^\top)^\top = \boldsymbol{\eta}^\top \mathbf{Q}_\mathrm{r} \boldsymbol{\eta} \geq 0$ because $\mathbf{Q}_\mathrm{r}$ is PSD; (ii) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}^\top)^\top = \eta_0^2(1, \frac{\boldsymbol{\eta}^\top}{\eta_0})\mathbf{Z}_k(1, \frac{\boldsymbol{\eta}^\top}{\eta_0})^\top \geq 0$ according to (A.22). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (A.23) shows that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ is in the null space of $\bar{\mathbf{B}}$ and thus cannot be represented by basis vectors in the space of $\bar{\mathbf{B}}$. Because $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$, we have $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ should be in the space of $\mathbf{B}$. That is, there exists $\mathbf{u}_k \in \mathbb{R}^{m_1}$ such that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}\mathbf{u}_k$ for any $k \in [K]$. Meanwhile, because $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}$, we have

$$\mathbf{B}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}^\top \mathbf{B}\mathbf{u}_k = \mathbf{u}_k,\ \forall k \in [K].$$

Finally, we obtain Problem (2.29) by replacing constraints (A.18b) with (A.24) and replacing $\mathbf{B}^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B})^\top(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ with $\mathbf{u}_k$. $\square$

A.4.3 Proof of Theorem 4

Consider $m_1 = K$. We construct a solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_{\mathrm{r}}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ of Problem (2.26) by setting $\mathbf{x}^\dagger = \mathbf{x}^*$, $s^\dagger = s^*$, $\hat{\boldsymbol{\lambda}}^\dagger = \hat{\boldsymbol{\lambda}}^*$, $\mathbf{q}^\dagger = \mathbf{q}' = \mathbf{V}\boldsymbol{\delta}$, $\mathbf{Q}_{\mathrm{r}}^\dagger = \mathbf{Y}_{11}$, $\mathbf{B}^\dagger = \mathbf{V}$, and $\hat{\mathbf{u}}_k^\dagger = \boldsymbol{\delta} + \boldsymbol{\nu}_k$ $(k \in [K])$.

First, we show this constructed solution is feasible to Problem (2.26). Clearly, this solution satisfies constraints (2.29d)–(2.29e). By the construction of the solution, for any $k \in [K]$, we further have

$$\begin{aligned}
\mathbf{q}^\dagger + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^\dagger - y_k(\mathbf{x}^\dagger)\right) &= \mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right) \\
&= \mathbf{V}\boldsymbol{\delta} + \mathbf{V}\boldsymbol{\nu}_k = \mathbf{B}^\dagger \hat{\mathbf{u}}_k^\dagger,
\end{aligned}$$

where the first equality holds by the construction of $\mathbf{q}^\dagger$, the second equality holds by (2.16), and the third equality holds by the construction of $\hat{\mathbf{u}}_k^\dagger$. Thus, this solution satisfies constraints (2.29c). Meanwhile, $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_K = \mathbf{I}_{m_1}$. It follows that $(\mathbf{B}^\dagger)^\top \mathbf{B}^\dagger = \mathbf{I}_{m_1}$.

In addition, from Problem (2.4), as $\mathbf{q}' = \mathbf{V}\boldsymbol{\delta}$ and $\mathbf{Q}' = \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top$, for any $k \in [K]$, we have

$$\begin{bmatrix}
S_k & \frac{1}{2}\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \\
\frac{1}{2}\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) & \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top
\end{bmatrix} \succeq 0,$$

which, by Schur complement, is equivalent to

$$\begin{aligned}
4S_k\left(\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top\right) &\succeq \left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)\left(\mathbf{V}\boldsymbol{\delta} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top \\
&= \left(\mathbf{V}\mathbf{u}_k^\dagger\right)\left(\mathbf{V}\mathbf{u}_k^\dagger\right)^\top, \qquad \text{(A.25)}
\end{aligned}$$

where the equality holds by (2.16) and the construction of $\hat{\mathbf{u}}_k^\dagger$. From (A.25), for any

$k \in [K]$, we have the following inequality holds by Lemma 2:

$$4S_k \left(\mathbf{V}^\top \mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top \mathbf{V}\right) \succeq \mathbf{V}^\top \left(\mathbf{V}\mathbf{u}_k^\dagger\right) \left(\mathbf{V}\mathbf{u}_k^\dagger\right)^\top \mathbf{V},$$

which is equivalent to

$$4S_k \mathbf{Y}_{11} \succeq \mathbf{u}_k^\dagger \mathbf{u}_k^{\dagger\top} \tag{A.26}$$

because $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_K$. By Schur complement, for any $k \in [K]$, (A.26) further becomes

$$\begin{bmatrix} S_k & \frac{1}{2}\mathbf{u}_k^{\dagger\top} \\ \frac{1}{2}\mathbf{u}_k^\dagger & \mathbf{Y}_{11} \end{bmatrix} \succeq 0,$$

which indicates that the constructed solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ also satisfies constraints (2.29b) and thus it is a feasible solution of Problem (2.26).

Second, we show this feasible solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ is an optimal solution of Problem (2.26). The objective value corresponding to this solution is

$$\begin{aligned} & s^\dagger + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^\dagger + \sqrt{\gamma_1} \left\|\mathbf{q}^\dagger\right\|_2 = s^* + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Y}_{11} + \sqrt{\gamma_1} \|\mathbf{q}'\|_2 \\ = & s^* + \gamma_2 \mathbf{I}_{m_1} \bullet (\mathbf{Y}_{11}\mathbf{V}^\top \mathbf{V}) + \sqrt{\gamma_1} \|\mathbf{q}'\|_2 = s^* + \gamma_2 \mathbf{I}_m \bullet (\mathbf{V}\mathbf{Y}_{11}\mathbf{V}^\top) + \sqrt{\gamma_1} \|\mathbf{q}'\|_2 \\ = & s^* + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1} \|\mathbf{q}'\|_2 = \Theta_\mathrm{M}(m), \end{aligned}$$

where the first equality holds by the construction of $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$, the second equality holds because $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_K$, the third equality holds by the cyclic property of a matrix's trace, and the fourth equality holds by the definition of $\mathbf{Q}'$ in Theorem 2. Therefore, the solution $(\mathbf{x}^\dagger, s^\dagger, \hat{\boldsymbol{\lambda}}^\dagger, \mathbf{q}^\dagger, \mathbf{Q}_\mathrm{r}^\dagger, \hat{\mathbf{u}}^\dagger, \mathbf{B}^\dagger)$ is an optimal solution of Problem (2.26).

Finally, when $m_1 > K$ and $m_1 \leq m$, we have $\Theta_\mathrm{U}(m_1) \geq \Theta_\mathrm{M}(m)$ by the conclusion (i) in Theorem 3 and $\Theta_\mathrm{U}(m_1) \leq \Theta_\mathrm{U}(K) = \Theta_\mathrm{M}(m)$ by the conclusion (ii) in Theorem 3. It follows that $\Theta_\mathrm{U}(m_1) = \Theta_\mathrm{M}(m)$. $\square$

A.4.4 Proof of Proposition 4

First, by Lemma 2, for any $\mathbf{B} \in \mathcal{B}_{m_1}$, we have

$$\mathbf{X} \preceq \mathbf{I}_m \Longrightarrow \mathbf{B}^\top \mathbf{X} \mathbf{B} \preceq \mathbf{B}^\top \mathbf{I}_m \mathbf{B} = \mathbf{I}_{m_1}.$$

Second, we perform eigenvalue decomposition on $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q} \in \mathbb{R}^{m \times m}$ is a matrix with orthonormal column vectors and $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ is a diagonal matrix. Without loss of generality, we assume that the diagonal elements of $\mathbf{\Lambda}$ are arranged in a nonincreasing order and let $\mathbf{\Lambda}_{m_1 \times m_1}$ represent the upper-left submatrix of $\mathbf{\Lambda}$.

Now we let $\mathbf{B} = \mathbf{Q}_{m \times m_1}$, where $\mathbf{Q}_{m \times m_1}$ is the left submatrix of $\mathbf{Q}$. Then we have $\mathbf{B} \in \mathcal{B}_{m_1}$ and

$$\begin{aligned}
\mathbf{B}^\top \mathbf{X} \mathbf{B} \preceq \mathbf{I}_{m_1} &\Longrightarrow \mathbf{B}^\top \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top \mathbf{B} \preceq \mathbf{I}_{m_1} \Longrightarrow \mathbf{Q}_{m \times m_1}^\top \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top \mathbf{Q}_{m \times m_1} \preceq \mathbf{I}_{m_1} \\
&\Longrightarrow [\mathbf{I}_{m_1}, \mathbf{0}_{m_1 \times (m-m_1)}]\mathbf{\Lambda}[\mathbf{I}_{m_1}, \mathbf{0}_{m_1 \times (m-m_1)}]^\top \preceq \mathbf{I}_{m_1} \\
&\Longrightarrow \mathbf{\Lambda}_{m_1 \times m_1} \preceq \mathbf{I}_{m_1} \Longrightarrow \mathbf{\Lambda} \preceq \mathbf{I}_m \Longrightarrow \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top \preceq \mathbf{Q}\mathbf{I}_m\mathbf{Q}^\top \Longrightarrow \mathbf{X} \preceq \mathbf{I}_m
\end{aligned}$$

where the first deduction holds by the eigenvalue decomposition of $\mathbf{X}$, the second deduction holds by the construction of $\mathbf{B}$, the third deduction holds because all the column vectors in $\mathbf{Q}$ are orthonormal, the fourth deduction holds by the definition of $\mathbf{\Lambda}_{m_1 \times m_1}$, the fifth deduction holds because $\mathrm{rank}(\mathbf{X}) \leq m_1$, the sixth deduction holds by Lemma 2. Thus, if $\mathbf{B}^\top \mathbf{X} \mathbf{B} \preceq \mathbf{I}_{m_1}$ for any $\mathbf{B} \in \mathcal{B}_{m_1}$, then we have $\mathbf{X} \preceq \mathbf{I}_m$. The proof is complete. $\square$

A.5 Supplement to Section 2.5

A.5.1 Proof of Theorem 5

By dualizing the inner maximization problem of Problem (2.30) and integrating it with the outer minimization operators, we first obtain the following formulation:

$$\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}'_{\mathrm{r}},\mathbf{Q}''_{\mathrm{r}}\\ \mathbf{B}_1,\mathbf{B}_2}} \quad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}'_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{A.27a}$$

$$\text{s.t.} \quad s \geq f\left(\mathbf{x}, \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) - \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_1 \mathbf{Q}'_{\mathrm{r}} \mathbf{B}_1^\top \boldsymbol{\xi}_{\mathrm{I}} - \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_2 \mathbf{Q}''_{\mathrm{r}} \mathbf{B}_2^\top \boldsymbol{\xi}_{\mathrm{I}} - \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}},\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}}, \tag{A.27b}$$

$$\mathbf{Q}'_{\mathrm{r}} \succeq 0,\ \mathbf{Q}''_{\mathrm{r}} \succeq 0,\ \mathbf{x} \in \mathcal{X},\ \mathbf{Q}'_{\mathrm{r}} \in \mathbb{R}^{m_1 \times m_1},\ \mathbf{Q}''_{\mathrm{r}} \in \mathbb{R}^{(K-m_1)\times(K-m_1)},\ \mathbf{q} \in \mathbb{R}^m, \tag{A.27c}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m \times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m \times (K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K. \tag{A.27d}$$

Next, we apply the strong duality theorem to constraints (A.27b). We define

$$g_k(\boldsymbol{\xi}_{\mathrm{I}}) = s + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_1 \mathbf{Q}'_{\mathrm{r}} \mathbf{B}_1^\top \boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\xi}_{\mathrm{I}}^\top \mathbf{B}_2 \mathbf{Q}''_{\mathrm{r}} \mathbf{B}_2^\top \boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top \boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right),\ \forall k \in [K].$$

As function $f(\mathbf{x}, \boldsymbol{\xi})$ is piecewise linear convex, we can reformulate (A.27b) as

$$g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0,\ \forall \boldsymbol{\xi}_{\mathrm{I}} \in \mathcal{S}_{\mathrm{I}},\ \forall k \in [K],$$

which is equivalent to

$$\min_{\mathbf{A}\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) \leq \mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}} \in \mathbb{R}^m} \quad g_k(\boldsymbol{\xi}_{\mathrm{I}}) \geq 0,\ \forall k \in [K]. \tag{A.28}$$

For any $k \in [K]$, the Lagrangian dual problem of $\min_{\mathbf{A}(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu})\leq\mathbf{b},\ \boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m}\ g_k(\boldsymbol{\xi}_{\mathrm{I}})$ is

$$\max_{\boldsymbol{\lambda}_k\geq 0}\ \min_{\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m}\ g_k(\boldsymbol{\xi}_{\mathrm{I}}) + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)-\mathbf{b}\right),$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^l$. Because there exists an interior point for the primal problem, the strong duality holds. Thus, constraints (A.28) are equivalent to

$$\max_{\boldsymbol{\lambda}_k\geq 0}\ \min_{\boldsymbol{\xi}_{\mathrm{I}}}\ g_k(\boldsymbol{\xi}_{\mathrm{I}}) + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)-\mathbf{b}\right)\geq 0,\ \forall k\in[K],$$

which are further equivalent to

$$\begin{aligned}\exists\boldsymbol{\lambda}_k\geq 0: s + \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_1\mathbf{Q}_{\mathrm{r}}'\mathbf{B}_1^\top\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\xi}_{\mathrm{I}}^\top\mathbf{B}_2\mathbf{Q}_{\mathrm{r}}''\mathbf{B}_2^\top\boldsymbol{\xi}_{\mathrm{I}} + \mathbf{q}^\top\boldsymbol{\xi}_{\mathrm{I}} - y_k^0(\mathbf{x}) - y_k(\mathbf{x})^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)&\\ + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}}+\boldsymbol{\mu}\right)-\mathbf{b}\right)\geq 0,\ \forall\boldsymbol{\xi}_{\mathrm{I}}\in\mathbb{R}^m,\ \forall k\in[K].&\end{aligned}\tag{A.29}$$

Note that $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$; that is, all the column vectors of $\mathbf{B}$ are orthogonal. We can then extend $\mathbf{B}$ to $[\mathbf{B},\bar{\mathbf{B}}]\in\mathbb{R}^{m\times m}$ with $\bar{\mathbf{B}}\in\mathbb{R}^{m\times(m-K)}$ such that all the column vectors of $[\mathbf{B},\bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$. Thus, we can always find $\boldsymbol{\xi}_1\in\mathbb{R}^{m_1}$, $\boldsymbol{\xi}_2\in\mathbb{R}^{K-m_1}$, and $\boldsymbol{\xi}_3\in\mathbb{R}^{m-K}$ such that

$$\boldsymbol{\xi}_{\mathrm{I}} = \mathbf{B}_1\boldsymbol{\xi}_1 + \mathbf{B}_2\boldsymbol{\xi}_2 + \bar{\mathbf{B}}\boldsymbol{\xi}_3.$$

It follows that constraints (A.29) become

$$\begin{aligned}\exists\boldsymbol{\lambda}_k\geq 0: s + \boldsymbol{\xi}_1^\top\mathbf{Q}_{\mathrm{r}}'\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2^\top\mathbf{Q}_{\mathrm{r}}''\boldsymbol{\xi}_2 + \mathbf{q}^\top\left(\mathbf{B}_1\boldsymbol{\xi}_1+\mathbf{B}_2\boldsymbol{\xi}_2+\bar{\mathbf{B}}\boldsymbol{\xi}_3\right) - y_k^0(\mathbf{x})&\\ - y_k(\mathbf{x})^\top\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}_1\boldsymbol{\xi}_1+\mathbf{B}_2\boldsymbol{\xi}_2+\bar{\mathbf{B}}\boldsymbol{\xi}_3\right)+\boldsymbol{\mu}\right)&\\ + \boldsymbol{\lambda}_k^\top\left(\mathbf{A}\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\left(\mathbf{B}_1\boldsymbol{\xi}_1+\mathbf{B}_2\boldsymbol{\xi}_2+\bar{\mathbf{B}}\boldsymbol{\xi}_3\right)+\boldsymbol{\mu}\right)-\mathbf{b}\right)\geq 0,&\\ \forall\boldsymbol{\xi}_1\in\mathbb{R}^{m_1},\boldsymbol{\xi}_2\in\mathbb{R}^{K-m_1},\boldsymbol{\xi}_3\in\mathbb{R}^{m-K},\ \forall k\in[K].&\end{aligned}\tag{A.30}$$

We further define

$$\mathbf{Z}_k = \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{h}_k')^\top & \frac{1}{2}(\mathbf{h}_k'')^\top \\ \frac{1}{2}\mathbf{h}_k' & \mathbf{Q}_{\mathrm{r}}' & \mathbf{0}_{m_1\times(K-m_1)} \\ \frac{1}{2}\mathbf{h}_k'' & \mathbf{0}_{(K-m_1)\times m_1} & \mathbf{Q}_{\mathrm{r}}'' \end{bmatrix}, \ \forall k \in [K],$$

where $\mathbf{h}_k' = \mathbf{B}_1^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ and $\mathbf{h}_k'' = \mathbf{B}_2^\top \mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2)^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ for any $k \in [K]$. It follows that

(A.30)

$$\begin{aligned}
&\iff \exists \boldsymbol{\lambda}_k \geq 0: \ \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top + \boldsymbol{\xi}_3^\top \left(\bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) \geq 0, \\
&\qquad\qquad \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \boldsymbol{\xi}_3 \in \mathbb{R}^{m-K}, \ \forall k \in [K]. \\
&\iff \exists \boldsymbol{\lambda}_k \geq 0: \ \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right) \mathbf{Z}_k \left(1, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right)^\top \geq 0, \ \forall \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\xi}_2 \in \mathbb{R}^{K-m_1}, \ \forall k \in [K]; \quad \text{(A.31)} \\
&\qquad\qquad \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0, \ \forall k \in [K]. \\
&\iff \exists \boldsymbol{\lambda}_k \geq 0: \ \mathbf{Z}_k \succeq 0, \ \bar{\mathbf{B}}^\top \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{B}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = 0, \ \forall k \in [K]. \\
&\iff \exists \boldsymbol{\lambda}_k \geq 0: \ \mathbf{Z}_k \succeq 0, \ \bar{\mathbf{B}}^\top \left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right)\right) = 0, \ \forall k \in [K]. \qquad \text{(A.32)} \\
&\iff \exists \boldsymbol{\lambda}_k \geq 0, \mathbf{u}_k' \in \mathbb{R}^{m_1}, \mathbf{u}_k'' \in \mathbb{R}^{K-m_1}: \\
&\qquad\qquad \mathbf{Z}_k \succeq 0, \ \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1\mathbf{u}_k' + \mathbf{B}_2\mathbf{u}_k'', \ \forall k \in [K]. \qquad \text{(A.33)}
\end{aligned}$$

The first equivalence holds due to the definition of $\mathbf{Z}_k$. For the third equivalence, clearly $\Longleftarrow$ follows from the definition of a PSD matrix. To prove $\Longrightarrow$, we consider two possible cases for any $(\eta_0 \in \mathbb{R}, \boldsymbol{\eta}_1^\top \in \mathbb{R}^{m_1}, \boldsymbol{\eta}_2^\top \in \mathbb{R}^{K-m_1})^\top \in \mathbb{R}^{K+1}$: (i) if $\eta_0 = 0$, then $(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top = \boldsymbol{\eta}_1^\top \mathbf{Q}_{\mathrm{r}}' \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^\top \mathbf{Q}_{\mathrm{r}}'' \boldsymbol{\eta}_2 \geq 0$ because $\mathbf{Q}_{\mathrm{r}}'$ and $\mathbf{Q}_{\mathrm{r}}''$ are PSD; (ii) if $\eta_0 \neq 0$, then we have $(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)\mathbf{Z}_k(\eta_0, \boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top = \eta_0^2(1, \frac{\boldsymbol{\eta}_1^\top}{\eta_0}, \frac{\boldsymbol{\eta}_2^\top}{\eta_0})\mathbf{Z}_k(1, \frac{\boldsymbol{\eta}_1^\top}{\eta_0}, \frac{\boldsymbol{\eta}_2^\top}{\eta_0})^\top \geq 0$ according to (A.31). Therefore, $\Longrightarrow$ holds. For the fifth equivalence, (A.32) shows that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ is in the null space of $\bar{\mathbf{B}}$ and thus cannot be represented by basis vectors in the space of $\bar{\mathbf{B}}$. Because $[\mathbf{B}, \bar{\mathbf{B}}]$ span the space of $\mathbb{R}^m$, we have $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x}))$ should be in the space of $\mathbf{B}$. That is, there exists $\mathbf{u}_k' \in \mathbb{R}^{m_1}$

and $\mathbf{u}''_k \in \mathbb{R}^{K-m_1}$ such that $\mathbf{q} + (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})) = \mathbf{B}_1\mathbf{u}'_k + \mathbf{B}_2\mathbf{u}''_k$ for any $k \in [K]$. Meanwhile, because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_K$, we have

$$\mathbf{h}'_k = \mathbf{B}_1^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_1\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1^\top\mathbf{B}_1\mathbf{u}'_k = \mathbf{u}'_k,\ \forall k \in [K],$$
$$\mathbf{h}''_k = \mathbf{B}_2^\top\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}_2\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_2^\top\mathbf{B}_2\mathbf{u}''_k = \mathbf{u}''_k,\ \forall k \in [K].$$

By replacing constraints (A.27b) with (A.33), we obtain the following problem:

$$\min_{\substack{\mathbf{x},s,\hat{\boldsymbol{\lambda}},\mathbf{q},\\ \mathbf{Q}'_{\mathrm{r}},\mathbf{Q}''_{\mathrm{r}},\hat{\mathbf{u}}',\hat{\mathbf{u}}'',\\ \mathbf{B}_1,\mathbf{B}_2}} \quad s + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}'_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{A.34a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}'_k)^\top & \frac{1}{2}(\mathbf{u}''_k)^\top \\ \frac{1}{2}\mathbf{u}'_k & \mathbf{Q}'_{\mathrm{r}} & \mathbf{0}_{m_1\times(K-m_1)} \\ \frac{1}{2}\mathbf{u}''_k & \mathbf{0}_{(K-m_1)\times m_1} & \mathbf{Q}''_{\mathrm{r}} \end{bmatrix} \succeq 0, \quad \forall k \in [K], \tag{A.34b}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1\mathbf{u}'_k + \mathbf{B}_2\mathbf{u}''_k,\ \forall k \in [K], \tag{A.34c}$$

$$\mathbf{x} \in \mathcal{X},\ \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}'_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{Q}''_{\mathrm{r}} \in \mathbb{R}^{(K-m_1)\times(K-m_1)}, \tag{A.34d}$$

$$\mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times(K-m_1)},\ [\mathbf{B}_1, \mathbf{B}_2]^\top[\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, \tag{A.34e}$$

$$\hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}^l_+,\ \forall k \in [K], \tag{A.34f}$$

$$\hat{\mathbf{u}}' = \{\mathbf{u}'_1, \ldots, \mathbf{u}'_K\},\ \mathbf{u}'_k \in \mathbb{R}^{m_1},\ \forall k \in [K], \tag{A.34g}$$

$$\hat{\mathbf{u}}'' = \{\mathbf{u}''_1, \ldots, \mathbf{u}''_K\},\ \mathbf{u}''_k \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. \tag{A.34h}$$

Note that the value of $\mathbf{Q}''_{\mathrm{r}}$ does not contribute to the objective function (A.34a). We can then let M be an arbitrarily large positive number and $\mathbf{Q}''_{\mathrm{r}} = M\mathbf{I}_{(K-m_1)\times(K-m_1)}$ be an optimal solution, by which constraints (A.34b) become

$$\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top\mathbf{b} - y_k(\mathbf{x})^\top\boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}'_k)^\top \\ \frac{1}{2}\mathbf{u}'_k & \mathbf{Q}'_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k \in [K]. \tag{A.35}$$

By replacing (A.34b) with (A.35), we obtain the formulation of Problem (2.31).

Based on the formulation of Problem (2.31), now we show that the three conclusions hold. Note that for any $\mathbf{B} \in \mathcal{B}_K = \{\mathbf{B} \in \mathbb{R}^{m\times K} \mid \mathbf{B}^\top\mathbf{B} = \mathbf{I}_K\}$, the optimal value of Problem (2.26), i.e., $\Theta_{\mathrm{U}}(K)$, reaches the optimal value of the original Problem (2.4), i.e., $\Theta_{\mathrm{M}}(m)$. We would like to show that by relaxing the constraints in Problem (2.26), we can obtain the exact formulation of Problem (2.31), thereby the three conclusions hold.

First, we rewrite constraints (2.29b)–(2.29c) in Problem (2.26) with $m_1 = K$ by dividing $\mathbf{B}$ into $[\mathbf{B}_1, \mathbf{B}_2]$ and $\mathbf{u}_k$ into $((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top)^\top$. Thus, we obtain the following formulation:

$$
\begin{aligned}
\min_{\substack{\mathbf{x}, s, \hat{\boldsymbol{\lambda}},\\ \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \hat{\mathbf{u}}', \hat{\mathbf{u}}'',\\ \mathbf{B}_1, \mathbf{B}_2}} \quad & s + \gamma_2 \mathbf{I}_K \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\, \|\mathbf{q}\|_2 && \text{(A.36a)}\\
\text{s.t.} \quad & \begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right) \\ \frac{1}{2}\left((\mathbf{u}_k')^\top, (\mathbf{u}_k'')^\top\right)^\top & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, \\
& \forall k \in [K], && \text{(A.36b)}\\
& \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_k - y_k(\mathbf{x})\right) = \mathbf{B}_1 \mathbf{u}_k' + \mathbf{B}_2 \mathbf{u}_k'',\ \forall k \in [K], && \text{(A.36c)}\\
& \mathbf{x} \in \mathcal{X},\ [\mathbf{B}_1, \mathbf{B}_2]^\top [\mathbf{B}_1, \mathbf{B}_2] = \mathbf{I}_K, && \text{(A.36d)}\\
& \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{K\times K},\ \mathbf{B}_1 \in \mathbb{R}^{m\times m_1}, \mathbf{B}_2 \in \mathbb{R}^{m\times (K-m_1)}, && \text{(A.36e)}\\
& \hat{\boldsymbol{\lambda}} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_K\},\ \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], && \text{(A.36f)}\\
& \hat{\mathbf{u}}' = \{\mathbf{u}_1', \ldots, \mathbf{u}_K'\},\ \mathbf{u}_k' \in \mathbb{R}^{m_1},\ \forall k \in [K], && \text{(A.36g)}\\
& \hat{\mathbf{u}}'' = \{\mathbf{u}_1'', \ldots, \mathbf{u}_K''\},\ \mathbf{u}_k'' \in \mathbb{R}^{K-m_1},\ \forall k \in [K]. && \text{(A.36h)}
\end{aligned}
$$

Second, we relax constraints (A.36b) into

$$
\begin{bmatrix} s - y_k^0(\mathbf{x}) - \boldsymbol{\lambda}_k^\top \mathbf{b} - y_k(\mathbf{x})^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}(\mathbf{u}_k')^\top \\ \frac{1}{2}\mathbf{u}_k' & \mathbf{Q}_{\mathrm{r}}' \end{bmatrix} \succeq 0,\ \forall k \in [K], \tag{A.37}
$$

where $\mathbf{Q}'_{\mathrm{r}} \in \mathbb{R}^{m_1 \times m_1}$ is the upper-left submatrix of $\mathbf{Q}_{\mathrm{r}}$. Note that if we use (A.37) to replace (A.36b), we obtain a relaxation and accordingly lower bound for Problem (A.36). In addition, we further reduce the optimal value of the relaxation by replacing $\mathbf{Q}_{\mathrm{r}}$ in the objective function (A.36a) with $\mathbf{Q}'_{\mathrm{r}}$. That is, we obtain a lower bound for the optimal value of Problem (2.26) with $m_1 = K$ (i.e., Problem (2.4)). After these two steps of relaxations, we obtain the exact formulation of Problem (2.31). Thus, we can conclude that Problem (2.31) is a relaxation of Problem (2.26) with $m_1 = K$. Therefore, by the conclusion in Theorem 4, we have

$$\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m).$$

That is, the conclusion (i) holds.

For the conclusion (ii): For any $0 \leq m_1 < m_2 \leq K$, we can follow the above two steps of relaxations to relax Problem (A.36) to the problem with the optimal value $\Theta_{\mathrm{L2}}(m_2)$, and based on this relaxed problem, we can further relax it to the problem with the optimal value $\Theta_{\mathrm{L2}}(m_1)$. Because all these problems are minimization problems, we have $\Theta_{\mathrm{L2}}(m_1) \leq \Theta_{\mathrm{L2}}(m_2)$.

For the conclusion (iii): When $m_1 = K$, Problem (2.31) becomes Problem (2.26) with $m_1 = K$. Thus, by the conclusion in Theorem 4, we have

$$\Theta_{\mathrm{L2}}(K) = \Theta_{\mathrm{U}}(K) = \Theta_{\mathrm{M}}(m).$$

A.6 Supplement to Section 2.6

A.6.1 Motivation to Derive ADMM Algorithms

Several techniques, including the McCormick envelopes and spatial branch-and-bound, can handle bilinear constraints and solve bilinear SDP problems like Problems (2.25), (2.26), and (2.31). However, to the best of our knowledge, the currently available bilinear SDP solvers, such as PENLAB and BMIBNB (Löfberg 2004), are not yet fully mature and only succeed on relatively small and simple problems. Therefore, we derive ADMM algorithms to solve the three approximations efficiently. To illustrate this, we consider the multiproduct newsvendor problem (2.39) in Section 2.7 as an example to compare our ADMM algorithms with benchmark solvers with default settings in solving this problem, where no time limit is given.

First, we use the Mosek solver to solve the original high-dimensional SDP problem (2.4) and report the optimality gap and computational time. Note that the Mosek solver uses the interior-point algorithm to solve an SDP problem and terminates when the relative gap between the primal and dual objective values is no greater than 10^{-9}. Second, we use the BMIBNB solver (the state-of-the-art bilinear SDP solver, to the best of our knowledge) to solve the low-dimensional bilinear SDP problem (2.26) with $m_1 = K$. Note that the BMIBNB solver uses the spatial branch-and-bound algorithm to solve a bilinear SDP problem. In the BMIBNB solver, we set "Mosek" as the lower bound solver and "fmincon" as the upper bound solver. The default termination condition of the BMIBNB solver is that the relative gap between the lower and upper bounds is no greater than 0.01 or the maximum number of nodes in the branch-and-bound tree is greater than 100. Third, we use our derived ADMM algorithms to solve Problems (2.25), (2.26), and (2.31), respectively. For each of the three problems, we follow the instructions in Section 2.6.2 to recover the lower or upper bounds in the following two steps: (i) We use the ADMM algorithm to solve the approximation problem and obtain a near-optimal dimensionality reduction matrix $\mathbf{B}^{\text{ADMM}}$; (ii) Given this $\mathbf{B}^{\text{ADMM}}$, we solve a low-dimensional SDP problem to recover the lower or upper bound for the original optimal value. Kindly note that the reported time for our lower or upper bounds reflects the combined time for both steps. Other numerical setups are detailed in Section 2.7.

Table A.2 shows the numerical results, where "Optimality Gap (%)" and "Time (secs)"

Table A.2: Performance of Bilinear SDP Solver on the Newsvendor Problem

	Size (m)-Instance Id	100-1	100-2	100-3	200-1	200-2	200-3
Mosek	Optimality Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00
	Time (secs)	12.32	11.34	13.36	342.00	432.57	368.04
BMIBNB	Optimality Gap (%)	4.31	3.89	3.65	4.25	3.83	3.85
	Time (secs)	3840.00	5552.10	6412.50	9885.50	9942.50	10643.00
ODR-LB	Gap1 (%)	0.07	0.07	0.09	0.00	0.00	0.00
	Time (secs)	0.58	0.72	0.83	0.42	0.91	0.88
ODR-RLB	Gap1 (%)	0.03	0.04	0.03	0.06	0.02	0.02
	Time (secs)	2.45	1.74	1.98	2.57	2.42	2.76
ODR-UB	Gap2 (%)	1.56	1.81	1.65	1.73	1.92	2.03
	Time (secs)	2.45	1.74	1.98	2.57	2.42	2.76

are reported by the corresponding solver, and "Gap1 (%)" (resp. "Gap2 (%)") represents the relative gap in percentage between a lower (resp. an upper) bound and the optimal value provided by the Mosek solver. Note that when $m > 200$, the BMIBNB solver fails to obtain any feasible solution within three hours, and thus, these cases are not reported in Table A.2. For the reported six instances, the Mosek solver can solve them to optimality, and the BMIBNB solver terminates when reaching the limit of the maximum number of nodes. These numerical results demonstrate that using the bilinear SDP solver to solve the low-dimensional bilinear SDP reformulation of the high-dimensional SDP leads to poor performance. It is not surprising. The Mosek solver employs the interior-point method to solve the high-dimensional SDP problem, leading to a polynomial-time algorithm. In contrast, the BMIBNB solver uses spatial branch-and-bound to solve the low-dimensional bilinear SDP problem, leading to an exponential-time algorithm. More importantly, the bilinear SDP problem is nonconvex, significantly increasing the computational challenge. Finally, these numerical results demonstrate that the ADMM algorithms bring significant improvement in terms of optimality gap and computational time.

A.6.2 The ADMM Algorithm for Problem (2.26)

Note that Algorithm 1 serves as a unified ADMM algorithm for all the three proposed approximations, i.e., Problems (2.25), (2.26), and (2.31). Here, we take Problem (2.26) as an example to demonstrate corresponding specific details for this problem. Recall that Problem (2.26) can be formulated as follows:

$$
\begin{aligned}
\Theta_{\mathrm{U}}(m_1) = \min_{\substack{\mathbf{B},\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]}} \quad & \phi(m_1, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}) && \text{(A.38a)}\\
\text{s.t.} \quad & \begin{bmatrix} \chi(k,\mathbf{x},s,\boldsymbol{\lambda}_k) & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0,\ \forall k\in[K], && \text{(A.38b)}\\
& \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \boldsymbol{\psi}(k,\mathbf{x},\boldsymbol{\lambda}_k) = \tilde{\mathbf{u}}_k,\ \forall k\in[K], && \text{(A.38c)}\\
& \mathbf{x}\in\mathcal{X},\ \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1},\ \mathbf{B}\in\mathbb{R}^{m\times m_1},\ \mathbf{q}\in\mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}}\in\mathbb{R}^{m_1\times m_1}, && \text{(A.38d)}\\
& \boldsymbol{\lambda}_k\in\mathbb{R}^l_+,\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \tilde{\mathbf{u}}_k\in\mathbb{R}^m,\ \forall k\in[K], && \text{(A.38e)}\\
& \tilde{\mathbf{u}}_k = \mathbf{B}\mathbf{u}_k,\ \forall k\in[K]. && \text{(A.38f)}
\end{aligned}
$$

where constraints (A.38f) and $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$ are bilinear constraints. We consider the following augmented Lagrangian problem for Problem (A.38):

$$
\begin{aligned}
\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K];\\ \mathbf{B};\ \boldsymbol{\beta}_k,\forall k\in[K]}} & \Bigg\{ s + \gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\left\|\mathbf{q}\right\|_2 + \sum_{k=1}^K \boldsymbol{\beta}_k^\top\left(\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right)\\
& + \sum_{k=1}^K \frac{\rho_k}{2}\left\|\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\right\|_2^2 \ \Bigg|\ \text{(A.38b)} - \text{(A.38e)} \Bigg\}, \qquad \text{(A.39)}
\end{aligned}
$$

where $\boldsymbol{\beta}_k \in \mathbb{R}^m$ ($\forall k\in[K]$) are Lagrangian multipliers and $\rho_k > 0$ ($\forall k\in[K]$) are the penalty parameters. Thus, we design Algorithm 2 to solve Problem (A.38).

In this algorithm, given $\mathbf{B}$ and $\boldsymbol{\beta}_k$ for any $k\in[K]$, Problem (A.39) becomes a low-dimensional (i.e., m_1+1) SDP problem. Given $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k\in[K])$, Problem (A.39) becomes a nonconvex optimization problem, while the following proposition shows that it has an analytical optimal solution. Thus, Algorithm 2 can be performed efficiently.

Algorithm 2 ADMM for Problem (A.38)

Initialize: $\mathbf{B}^0, \boldsymbol{\beta}_k^0, \forall k \in [K]$
Repeat: update $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \mathbf{B}$ and $\boldsymbol{\beta}_k(\forall k \in [K])$ alternatingly by

Given $\mathbf{B}^i$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (A.39) to obtain the optimal solution $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$;

Given $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ and $\boldsymbol{\beta}_k^i$ for any $k \in [K]$, solve Problem (A.39) to obtain the optimal solution $\mathbf{B}^{i+1}$;

$\boldsymbol{\beta}_k^{i+1} = \boldsymbol{\beta}_k^i + \rho_k^i \left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1}\mathbf{u}_k^{i+1}\right), \ \forall k \in [K]$;
Until Convergence.

Proposition 12. *Given* $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$*, Problem* (A.39) *has an optimal solution* $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$*, where* $\sum_{k=1}^K (\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho_k \tilde{\mathbf{u}}_k \mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$ *for* $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times m_1}$*,* $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{m_1 \times m_1}$*, and* $\tilde{\mathbf{V}} \in \mathbb{R}^{m_1 \times m_1}$ *by the singular value decomposition (SVD).*

We further analyze the convergence property of Algorithm 2 to ensure the dimensionality reduction solution $\mathbf{B}$ returned by this algorithm is near-optimal, i.e., a theoretical guarantee. First, the following lemma holds.

Lemma 5. *Given* $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$*, we have* $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ *(an optimal solution of Problem* (A.39)*) is also an optimal solution of the following convex optimization problem:*

$$\max_{\mathbf{B} \in \mathbb{R}^{m \times m_1}} \left\{ \sum_{k=1}^K \left(\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho \tilde{\mathbf{u}}_k \mathbf{u}_k^\top\right) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top \mathbf{B} \preceq \mathbf{I}_{m_1} \right\}. \tag{A.40}$$

We then present the convergence properties of our proposed ADMM algorithm. We let

$$\begin{aligned} &\mathcal{L}\left(\mathbf{B}, (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), (\boldsymbol{\beta}_k, \forall k \in [K])\right) = \\ &\qquad s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\|\mathbf{q}\|_2 + \sum_{k=1}^K \boldsymbol{\beta}_k^\top (\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k) + \sum_{k=1}^K \frac{\rho_k}{2}\|\tilde{\mathbf{u}}_k - \mathbf{B}\mathbf{u}_k\|_2^2, \end{aligned}$$

and continue to adopt Assumption 3. We state the convergence theorem of Algorithm 2 as follows.

Theorem 10. *Let* $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_{\mathrm{r}}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *be any accumulation point of the sequence* $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_{\mathrm{r}}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$ *generated by Algorithm 2. Then,* $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_{\mathrm{r}}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ *satisfies the first-order stationary conditions of Problem* (A.38).

Proof. Proof. Let $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_{\mathrm{r}}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ be an accumulation point of the sequence $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_{\mathrm{r}}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$. Then, there exists a subsequence $\{\mathbf{B}^i, \mathbf{x}^i, s^i, \mathbf{q}^i, \mathbf{Q}_{\mathrm{r}}^i, \boldsymbol{\lambda}_k^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}_{i\in\mathcal{I}}$ that converges to $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_{\mathrm{r}}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$.

First, note that $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ is the optimal solution of the convex problem

$$\min_{\substack{\mathbf{x},s,\mathbf{q},\mathbf{Q}_{\mathrm{r}},\\ \boldsymbol{\lambda}_k,\tilde{\mathbf{u}}_k,\mathbf{u}_k,\forall k\in[K]}} \left\{\mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), (\boldsymbol{\beta}_k^i, \forall k \in [K])\right) \;\middle|\; (\text{A.38b}) - (\text{A.38e})\right\}. \tag{A.41}$$

We show that there exists an interior point in the feasible region of Problem (A.41), by which the KKT conditions are first-order necessary conditions for the optimal solution of Problem (A.41). Specifically, by letting $\mathbf{x}'$ be an interior point in $\mathcal{X}$ (by Assumption 2), we can set $\hat{\boldsymbol{\lambda}}' = \{\mathbf{1}_l, \ldots, \mathbf{1}_l\}$, $s' = \sum_{k=1}^K \left|y_k^0(\mathbf{x}') + \mathbf{1}_l^\top \mathbf{b} + y_k(\mathbf{x}')^\top \boldsymbol{\mu} - \mathbf{1}_l^\top \mathbf{A}\boldsymbol{\mu}\right| + 1$, $\mathbf{q}' = \mathbf{0}_m$, $\tilde{\mathbf{u}}_k = (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^\top(\mathbf{A}^\top \mathbf{1}_l - y_k(\mathbf{x}'))$, $\mathbf{u}_k = \mathbf{0}_{m_1}$, and $\mathbf{Q}_{\mathrm{r}}' = \mathbf{I}_{m_1}$. Clearly, $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}_{\mathrm{r}}', \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])$ is an interior point in the feasible region of Problem (A.41). Therefore, the following first-order stationary conditions hold; that is, there exists $\begin{bmatrix} t_k^i & (\mathbf{p}_k^i)^\top \\ \mathbf{p}_k^i & \mathbf{P}_k^i \end{bmatrix} \succeq 0$, $\boldsymbol{\pi}_k^i \geq 0$, $\boldsymbol{\eta}_k^i \in \mathbb{R}^m$, and $\mathbf{Z}^i \succeq 0$ such that

$$1 - \sum_{k=1}^K t_k^i = 0, \tag{A.42a}$$

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k^i = \mathbf{0}_{m_1 \times m_1}, \tag{A.42b}$$

$$\frac{\sqrt{\gamma_1}(\mathbf{q}^{i+1})^\top}{\|\mathbf{q}^{i+1}\|_2} + \sum_{k=1}^{K} (\boldsymbol{\eta}_k^i)^\top = 0, \tag{A.42c}$$

$$t_k^i (\mathbf{b} - \mathbf{A}\boldsymbol{\mu})^\top + (\boldsymbol{\eta}_k^i)^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top - (\boldsymbol{\pi}_k^i)^\top = 0, \ \forall k \in [K], \tag{A.42d}$$

$$-(\mathbf{p}_k^i)^\top - (\boldsymbol{\beta}_k^i)^\top \mathbf{B}^i - \rho_k^i (\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^i \mathbf{u}_k^{i+1}) \mathbf{B}^i = 0, \ \forall k \in [K], \tag{A.42e}$$

$$-(\boldsymbol{\eta}_k^i)^\top + (\boldsymbol{\beta}_k^i)^\top + \rho_k^i (\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^i \mathbf{u}_k^{i+1}) = 0, \ \forall k \in [K], \tag{A.42f}$$

$$\sum_{k=1}^{K} \left(t_k^i \mathbf{w}_k^0 + \left(t_k^i \boldsymbol{\mu}^\top + (\boldsymbol{\eta}_k^i)^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \right) \mathbf{W}_k \right) - \sum_{i'=1}^{\tau} \sum_{j'=1}^{\tau} z_{i'j'}^i \mathbf{a}_{i'j'} = 0. \tag{A.42g}$$

By (A.42a) and $t_k^i \geq 0$ ($\forall k \in [K]$), we have that $\{t_k^i\}_{i \in \mathcal{I}}$ ($\forall k \in [K]$) are bounded. Because every bounded sequence has a convergent subsequence, without loss of generality, we can assume that $t_k^i \to t_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$). Taking limits in (A.42a) for $i \in \mathcal{I}$, we have

$$1 - \sum_{k=1}^{K} t_k^* = 0. \tag{A.43}$$

Similarly, by (A.42b), we have $\mathbf{P}_k^i \to \mathbf{P}_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$) and

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k^* = \mathbf{0}_{m_1 \times m_1}. \tag{A.44}$$

Taking limits in (A.42e) and (A.42f) for $i \in \mathcal{I}$, by (2.35)and (2.36), we have that $\mathbf{p}_k^i \to \mathbf{p}_k^*$, $\boldsymbol{\eta}_k^i \to \boldsymbol{\eta}_k^*$ ($i \to \infty, i \in \mathcal{I}, \forall k \in [K]$), and

$$-(\mathbf{p}_k^*)^\top - (\boldsymbol{\beta}_k^*)^\top \mathbf{B}^* = 0, \ \forall k \in [K], \tag{A.45}$$

$$-(\boldsymbol{\eta}_k^*)^\top + (\boldsymbol{\beta}_k^*)^\top = 0, \ \forall k \in [K]. \tag{A.46}$$

Taking limits in (A.42d) and (A.42g) for $i \in \mathcal{I}$, we have that $\mathbf{Z}^i \to \mathbf{Z}^*$, $\boldsymbol{\pi}_k^i \to \boldsymbol{\pi}_k^*$

($\forall k \in [K]$), and

$$t_k^* (\mathbf{b} - \mathbf{A}\boldsymbol{\mu})^\top + (\boldsymbol{\eta}_k^*)^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top - (\boldsymbol{\pi}_k^*)^\top = 0,\ \forall k \in [K], \tag{A.47}$$

$$\sum_{k=1}^{K} \left(t_k^* \mathbf{w}_k^0 + \left(t_k^* \boldsymbol{\mu}^\top + (\boldsymbol{\eta}_k^*)^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \right) \mathbf{W}_k \right) - \sum_{i'=1}^{\tau} \sum_{j'=1}^{\tau} z_{i'j'}^* \mathbf{a}_{i'j'} = 0. \tag{A.48}$$

Taking limits in (A.42c) for $i \in \mathcal{I}$, we have

$$\frac{\sqrt{\gamma_1}(\mathbf{q}^*)^\top}{\|\mathbf{q}^*\|_2} + \sum_{k=1}^{K} (\boldsymbol{\eta}_k^*)^\top = 0. \tag{A.49}$$

Next, note that $\mathbf{B}^{i+1}$ is the optimal solution of the nonconvex problem

$$\min_{\mathbf{B}} \left\{ \mathcal{L} \left(\mathbf{B}, (\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_{\mathrm{r}}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}, \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right) \right) \;\middle|\; \text{(A.38b)} - \text{(A.38e)} \right\}.$$

In Proposition 12, we give an analytical optimal solution of $\mathbf{B}^{i+1}$. However, there is no optimality condition for this $\mathbf{B}^{i+1}$. By Lemma 5, we have that $\mathbf{B}^{i+1}$ is an optimal solution of the convex problem (A.40). Clearly, $\mathbf{B} = \mathbf{0}_{m \times m_1}$ is an interior point of Problem (A.40). Therefore, $\mathbf{B}^{i+1}$ satisfies the first-order stationary conditions of Problem (A.40); that is, there exists $\mathbf{C}^{i+1} \succeq 0$ such that

$$-\sum_{k=1}^{K} \boldsymbol{\beta}_k^i (\mathbf{u}_k^{i+1})^\top + \mathbf{B}^{i+1} (\mathbf{C}^{i+1})^\top - \rho_k^i \sum_{k=1}^{K} \left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1} \mathbf{u}_k^i \right) (\mathbf{u}_k^i)^\top = 0. \tag{A.50}$$

Taking limits in (A.50) for $i \in \mathcal{I}$, we have that $\mathbf{B}^i (\mathbf{C}^i)^\top \to \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top$ ($i \to \infty, \forall i \in \mathcal{I}$). Because $(\mathbf{B}^i)^\top \mathbf{B}^i = \mathbf{I}_{m_1}$ and $\mathbf{B}^i \to \mathbf{B}^*$ ($i \to \infty, \forall i \in \mathcal{I}$), it follows that

$$\begin{aligned}
& (\mathbf{B}^i)^\top \mathbf{B}^i (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I}, \\
\Longrightarrow\ & (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I}, \\
\Longrightarrow\ & (\mathbf{C}^i)^\top \to (\mathbf{B}^*)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},
\end{aligned}$$

$$\Longrightarrow \mathbf{C}^i \to \mathbf{C}^* := \left(\sum_{k=1}^{K} \mathbf{u}_k^* (\boldsymbol{\beta}_k^*)^\top\right) \mathbf{B}^*,\ i \to \infty, \forall i \in \mathcal{I}.$$

Thus, taking limits in (A.50) for $i \in \mathcal{I}$, we have

$$-\sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top + \mathbf{B}^* (\mathbf{C}^*)^\top = 0. \tag{A.51}$$

Finally, combining (A.43)–(A.49) and (A.51), we have that $(\mathbf{B}^*, \mathbf{x}^*, s^*, \mathbf{q}^*, \mathbf{Q}_{\mathrm{r}}^*, \boldsymbol{\lambda}_k^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ satisfies the stationary conditions of Problem (A.38); that is, there exists $\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0$, $\boldsymbol{\eta}_k$, $\boldsymbol{\omega}_k$, $\boldsymbol{\pi}_k \geq 0$, $\mathbf{C}$, and $\mathbf{Z} \succeq 0$ such that

$$1 - \sum_{k=1}^{K} t_k = 0, \tag{A.52a}$$

$$\gamma_2 \mathbf{I}_{m_1} - \sum_{k=1}^{K} \mathbf{P}_k = \mathbf{0}_{m_1 \times m_1}, \tag{A.52b}$$

$$\frac{\sqrt{\gamma_1} (\mathbf{q}^*)^\top}{\|\mathbf{q}^*\|_2} + \sum_{k=1}^{K} \boldsymbol{\eta}_k^\top = 0, \tag{A.52c}$$

$$t_k (\mathbf{b} - \mathbf{A}\boldsymbol{\mu})^\top + \boldsymbol{\eta}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top - \boldsymbol{\pi}_k^\top = 0,\ \forall k \in [K], \tag{A.52d}$$

$$-\mathbf{p}_k^\top - \boldsymbol{\omega}_k^\top \mathbf{B}^* = 0,\ \forall k \in [K], \tag{A.52e}$$

$$-\boldsymbol{\eta}_k^\top + \boldsymbol{\omega}_k^\top = 0,\ \forall k \in [K], \tag{A.52f}$$

$$-\sum_{k=1}^{K} \boldsymbol{\omega}_k (\mathbf{u}_k^*)^\top + \mathbf{B}^* \mathbf{C}^\top = 0, \tag{A.52g}$$

$$\sum_{k=1}^{K} \left(t_k \mathbf{w}_k^0 + \left(t_k \boldsymbol{\mu}^\top + \boldsymbol{\eta}_k^\top \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \right) \mathbf{W}_k \right) - \sum_{i=1}^{\tau} \sum_{j=1}^{\tau} z_{ij} \mathbf{a}_{ij} = 0. \tag{A.52h}$$

This completes the proof. □

A.6.3 Proof of Propositions 5 and 12

In Proposition 5, given $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, we can omit the constant in the objective function of Problem (2.33) and rewrite this problem as follows:

$$\min_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{ \sum_{k=1}^{K} -\boldsymbol{\beta}_k^\top \mathbf{B}\mathbf{u}_k + \sum_{k=1}^{K} \left(-\rho_k \tilde{\mathbf{u}}_k^\top \mathbf{B}\mathbf{u}_k + \frac{\rho_k}{2} \mathbf{u}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{u}_k \right) \;\middle|\; \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1} \right\}. \quad \text{(A.53)}$$

In Proposition 12, given $(\mathbf{x}, s, \mathbf{q}, \mathbf{Q}_\mathrm{r}, \boldsymbol{\lambda}_k, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \boldsymbol{\beta}_k, \forall k \in [K])$, we can also omit the constant in the objective function of Problem (A.39) and rewrite this problem as Problem (A.53).

By $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, the term $\mathbf{u}_k^\top \mathbf{B}^\top \mathbf{B}\mathbf{u}_k$ is also a constant. Because $\boldsymbol{\beta}_k^\top \mathbf{B}\mathbf{u}_k = (\boldsymbol{\beta}_k \mathbf{u}_k^\top) \bullet \mathbf{B}$ and $\tilde{\mathbf{u}}_k^\top \mathbf{B}\mathbf{u}_k = (\tilde{\mathbf{u}}_k \mathbf{u}_k^\top) \bullet \mathbf{B}$, we can further rewrite Problem (A.53) as follows:

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{ \sum_{k=1}^{K} (\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho_k \tilde{\mathbf{u}}_k \mathbf{u}_k^\top) \bullet \mathbf{B} \;\middle|\; \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1} \right\}.$$

By the SVD, i.e., $\sum_{k=1}^{K} (\boldsymbol{\beta}_k \mathbf{u}_k^\top + \rho_k \tilde{\mathbf{u}}_k \mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$, we have

$$\begin{aligned}
\mathbf{B}^* &= \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \left(\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\right) \bullet \mathbf{B} = \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} tr\left(\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top \mathbf{B}^\top\right) \\
&= \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} tr\left(\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top \mathbf{B}^\top \tilde{\mathbf{U}}\right) = \arg\max_{\mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1}} \tilde{\boldsymbol{\Sigma}} \bullet \left(\tilde{\mathbf{U}}^\top \mathbf{B}\tilde{\mathbf{V}}\right),
\end{aligned}$$

where the second and fourth equalities hold by the definition of a matrix's trace and the third equality holds by the cyclic property of a matrix's trace. Eldén and Park (1999) show that $\mathbf{B}^* = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^\top$ is an optimal solution. $\square$

A.6.4 Proof of Lemmas 3 and 5

By the SVD in Propositions 5 and 12, we have

$$\sum_{k=1}^{K}(\boldsymbol{\beta}_k\mathbf{u}_k^\top + \rho_k\tilde{\mathbf{u}}_k\mathbf{u}_k^\top) = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top.$$

We construct $\bar{\mathbf{U}} \in \mathbb{R}^{m\times m}$ by adding $m-m_1$ orthonormal columns to $\tilde{\mathbf{U}} \in \mathbb{R}^{m\times m_1}$ such that $\bar{\mathbf{U}}^\top\bar{\mathbf{U}} = \mathbf{I}_m$, and add $m-m_1$ zero rows to $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{m_1\times m_1}$ to construct $\bar{\boldsymbol{\Sigma}} = [\tilde{\boldsymbol{\Sigma}}; \mathbf{0}_{(m-m_1)\times m_1}] \in \mathbb{R}^{m\times m_1}$. It follows that $\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top = \bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top$.

Meanwhile, by the cyclic property of a matrix's trace, we have

$$\left(\bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\right) \bullet \mathbf{B} = tr\left(\bar{\mathbf{U}}\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\right) = tr\left(\bar{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^\top\mathbf{B}^\top\bar{\mathbf{U}}\right) = \bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right).$$

It follows that Problem (A.40) is equivalent to

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{\bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right) \,\middle|\, \mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}\right\}. \tag{A.54}$$

Note that we have $\bar{\mathbf{U}}\bar{\mathbf{U}}^\top = \mathbf{I}_m$ and $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top = \mathbf{I}_{m_1}$ by the SVD and the construction of $\bar{\mathbf{U}}$. We then have

$$\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1} \iff \mathbf{B}^\top\bar{\mathbf{U}}\bar{\mathbf{U}}^\top\mathbf{B} \preceq \mathbf{I}_{m_1} \iff \tilde{\mathbf{V}}^\top\mathbf{B}^\top\bar{\mathbf{U}}\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}} \preceq \mathbf{I}_{m_1},$$

where the first equivalence holds by $\bar{\mathbf{U}}\bar{\mathbf{U}}^\top = \mathbf{I}_m$ and the second equivalence holds by $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\top = \mathbf{I}_{m_1}$ and Lemma 2.

Therefore, Problem (A.54) is equivalent to

$$\max_{\mathbf{B}\in\mathbb{R}^{m\times m_1}} \left\{\bar{\boldsymbol{\Sigma}} \bullet \left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right) \,\middle|\, \left(\tilde{\mathbf{V}}^\top\mathbf{B}^\top\bar{\mathbf{U}}\right)\left(\bar{\mathbf{U}}^\top\mathbf{B}\tilde{\mathbf{V}}\right) \preceq \mathbf{I}_{m_1}\right\}. \tag{A.55}$$

Because $\bar{\mathbf{\Sigma}} \in \mathbb{R}^{m \times m_1}$ is a rectangular diagonal matrix with non-negative numbers on the diagonal, we have that $\bar{\mathbf{U}}^\top \mathbf{B} \tilde{\mathbf{V}} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}]$ will lead to the optimal value. Note that when we choose $\mathbf{B} = \tilde{\mathbf{U}} \tilde{\mathbf{V}}^\top$, we have

$$\bar{\mathbf{U}}^\top \mathbf{B} \tilde{\mathbf{V}} = \bar{\mathbf{U}}^\top \tilde{\mathbf{U}} \tilde{\mathbf{V}}^\top \tilde{\mathbf{V}} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}] \mathbf{I}_{m_1} = [\mathbf{I}_{m_1}; \mathbf{0}_{(m-m_1)\times m_1}].$$

Therefore, $\mathbf{B}^* = \tilde{\mathbf{U}} \tilde{\mathbf{V}}^\top$ is an optimal solution of Problem (A.55), i.e., Problem (A.40). □

A.6.5 Proof of Theorem 6

Let $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ be an accumulation point of the sequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}$. Then, there exists a subsequence $\{\mathbf{B}^i, \mathbf{x}^i, \tilde{\mathbf{u}}_k^i, \mathbf{u}_k^i, \forall k \in [K]\}_{i \in \mathcal{I}}$ that converges to $(\mathbf{B}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$.

First, note that $(\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}$ is the optimal solution of the convex problem

$$\min_{\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]} \left\{ \mathcal{L}\left(\mathbf{B}^i, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K]), \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \;\middle|\; (2.32\text{b}) - (2.32\text{c}) \right\}. \quad \text{(A.56)}$$

Because there exists an interior point in the feasible region of Problem (A.56), by which the KKT conditions are first-order necessary conditions for the optimal solution of Problem (A.56). Specifically, the following first-order stationary conditions hold; that is,

$$-\nabla g(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^i)^\top \boldsymbol{\beta}_1^i + \rho_1^i (\mathbf{B}^i)^\top \left(\tilde{\mathbf{u}}_1^{i+1} - \mathbf{B}^i \mathbf{u}_1^{i+1}\right) \\ \vdots \\ (\mathbf{B}^i)^\top \boldsymbol{\beta}_K^i + \rho_K^i (\mathbf{B}^i)^\top \left(\tilde{\mathbf{u}}_K^{i+1} - \mathbf{B}^i \mathbf{u}_K^{i+1}\right) \\ -\boldsymbol{\beta}_1^i - \rho_1^i \left(\tilde{\mathbf{u}}_1^{i+1} - \mathbf{B}^i \mathbf{u}_1^{i+1}\right) \\ \vdots \\ -\boldsymbol{\beta}_K^i - \rho_K^i \left(\tilde{\mathbf{u}}_K^{i+1} - \mathbf{B}^i \mathbf{u}_K^{i+1}\right) \end{bmatrix}$$

$$\in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K]), \tag{A.57}$$

where $\mathcal{N}_{\mathcal{U}}(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K])$ is the normal cone of $\mathcal{U}$ at $(\mathbf{x}^{i+1}, \mathbf{u}_k^{i+1}, \tilde{\mathbf{u}}_k^{i+1}, \forall k \in [K])$. Taking limits in (A.57) for $i \in \mathcal{I}$, we have

$$-\nabla g(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^*)^\top \boldsymbol{\beta}_1^* \\ \vdots \\ (\mathbf{B}^*)^\top \boldsymbol{\beta}_K^* \\ -\boldsymbol{\beta}_1^* \\ \vdots \\ -\boldsymbol{\beta}_K^* \end{bmatrix} \in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]). \tag{A.58}$$

Next, note that $\mathbf{B}^{i+1}$ is the optimal solution of the nonconvex problem

$$\min_{\mathbf{B}} \left\{ \mathcal{L}\left(\mathbf{B}, (\mathbf{x}, \tilde{\mathbf{u}}_k, \mathbf{u}_k, \forall k \in [K])^{i+1}, \left(\boldsymbol{\beta}_k^i, \forall k \in [K]\right)\right) \;\middle|\; (2.32\text{b}) - (2.32\text{c}) \right\}.$$

In Proposition 5, we give an analytical optimal solution of $\mathbf{B}^{i+1}$. However, there is no optimality condition for this $\mathbf{B}^{i+1}$. By Lemma 3, we have that $\mathbf{B}^{i+1}$ is an optimal solution of the convex problem (2.34). Clearly, $\mathbf{B} = \mathbf{0}_{m \times m_1}$ is an interior point of Problem (2.34). Therefore, $\mathbf{B}^{i+1}$ satisfies the first-order stationary conditions of Problem (2.34); that is, there exists $\mathbf{C}^{i+1} \succeq 0$ such that

$$-\sum_{k=1}^{K} \boldsymbol{\beta}_k^i (\mathbf{u}_k^{i+1})^\top + \mathbf{B}^{i+1} (\mathbf{C}^{i+1})^\top - \rho_k^i \sum_{k=1}^{K} \left(\tilde{\mathbf{u}}_k^{i+1} - \mathbf{B}^{i+1} \mathbf{u}_k^i\right) (\mathbf{u}_k^i)^\top = 0. \tag{A.59}$$

Taking limits in (A.59) for $i \in \mathcal{I}$, we have that $\mathbf{B}^i (\mathbf{C}^i)^\top \to \sum_{k=1}^K \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top$ $(i \to \infty, \forall i \in \mathcal{I})$. Because $(\mathbf{B}^i)^\top \mathbf{B}^i = \mathbf{I}_{m_1}$ and $\mathbf{B}^i \to \mathbf{B}^*$ $(i \to \infty, \forall i \in \mathcal{I})$, it follows that

$$(\mathbf{B}^i)^\top \mathbf{B}^i (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^{K} \boldsymbol{\beta}_k^* (\mathbf{u}_k^*)^\top, \;\; i \to \infty, \forall i \in \mathcal{I},$$

$$\Longrightarrow (\mathbf{C}^i)^\top \to (\mathbf{B}^i)^\top \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},$$
$$\Longrightarrow (\mathbf{C}^i)^\top \to (\mathbf{B}^*)^\top \sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top,\ i \to \infty, \forall i \in \mathcal{I},$$
$$\Longrightarrow \mathbf{C}^i \to \mathbf{C}^* := \left(\sum_{k=1}^K \mathbf{u}_k^*(\boldsymbol{\beta}_k^*)^\top\right)\mathbf{B}^*,\ i \to \infty, \forall i \in \mathcal{I}.$$

Thus, taking limits in (A.59) for $i \in \mathcal{I}$, we have

$$-\sum_{k=1}^K \boldsymbol{\beta}_k^*(\mathbf{u}_k^*)^\top + \mathbf{B}^*(\mathbf{C}^*)^\top = 0. \tag{A.60}$$

Finally, combining (A.58) and (A.60), we have that $(\mathbf{B}^*, \mathbf{x}^*, \tilde{\mathbf{u}}_k^*, \mathbf{u}_k^*, \forall k \in [K])$ satisfies the stationary conditions of Problem (2.33); that is, there exist $\boldsymbol{\omega}_k$ and $\mathbf{C}$ such that

$$-\nabla g(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K]) + \begin{bmatrix} \mathbf{0} \\ (\mathbf{B}^*)^\top \boldsymbol{\omega}_1 \\ \vdots \\ (\mathbf{B}^*)^\top \boldsymbol{\omega}_K \\ -\boldsymbol{\omega}_1 \\ \vdots \\ -\boldsymbol{\omega}_K \end{bmatrix} \in \mathcal{N}_{\mathcal{U}}(\mathbf{x}^*, \mathbf{u}_k^*, \tilde{\mathbf{u}}_k^*, \forall k \in [K])$$
$$-\sum_{k=1}^K \boldsymbol{\omega}_k(\mathbf{u}_k^*)^\top + \mathbf{B}^*\mathbf{C}^\top = 0 \tag{A.61}$$

This completes the proof. □

A.6.6 Proof of Proposition 6

First, we have $\Theta_{\mathrm{M}}(m) \geq \Theta_{\mathrm{L}}(m_1) \geq \underline{\Theta}(m_1, \mathbf{B}') = s^* + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}}^* + \sqrt{\gamma_1}\|\mathbf{q}_{\mathrm{r}}^*\|_2$, where the first inequality holds by conclusion (i) of Theorem 1 and the second inequality holds because $\mathbf{B}'$ is a feasible solution of Problem (2.12) and this problem is a maximization problem.

Next, we would like to construct a feasible solution $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ of Problem (2.4). We set $\mathbf{x}' = \mathbf{x}^*$, $\hat{\boldsymbol{\lambda}}' = \hat{\boldsymbol{\lambda}}^*$, $s' = s^* + s_0$, $\mathbf{q}' = \mathbf{B}'\mathbf{q}_\mathrm{r}^*$, and $\mathbf{Q}' = \mathbf{B}'\mathbf{Q}_\mathrm{r}^*(\mathbf{B}')^\top + \mathbf{Q}_0$, where $s_0 \geq 0$ and $\mathbf{Q}_0 \succeq 0$ and their values will be decided later. Clearly, this solution satisfies constraints (2.4c). For this solution to satisfy constraints (2.4b), the values s_0 and $\mathbf{Q}_0$ should satisfy

$$(S_k + s_0)\left(\mathbf{B}'\mathbf{Q}_\mathrm{r}^*(\mathbf{B}')^\top + \mathbf{Q}_0\right) \succeq \frac{1}{4}\left(\mathbf{B}'\mathbf{q}_\mathrm{r}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right) \times \left(\mathbf{B}'\mathbf{q}_\mathrm{r}^* + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top\boldsymbol{\lambda}_k^* - y_k(\mathbf{x}^*)\right)\right)^\top = \frac{1}{4}\mathbf{M}_k, \ \forall k \in [K]. \tag{A.62}$$

Note that, if $(S + s_0)\mathbf{Q}_0 \succeq (1/4)\mathbf{M}_k$ for any $k \in [K]$, then (A.62) holds. By the definition of $\mathbf{M}_k$, we have $\mathbf{M}_k \succeq 0$ for any $k \in [K]$. Therefore, for any $s_0 \geq 0$, we can construct

$$\mathbf{Q}_0 = \sum_{k=1}^K \frac{1}{4(S + s_0)}\mathbf{M}_k$$

such that (A.62) holds and hence $(\mathbf{x}', s', \hat{\boldsymbol{\lambda}}', \mathbf{q}', \mathbf{Q}')$ is a feasible solution of Problem (2.4). The objective value (denoted by Θ'_M) with respect to this constructed solution is

$$\begin{aligned} s' + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}' + \sqrt{\gamma_1}\,\|\mathbf{q}'\|_2 &= s^* + s_0 + \gamma_2\mathbf{I}_m \bullet \mathbf{B}'\mathbf{Q}_\mathrm{r}^*(\mathbf{B}')^\top + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}_0 + \sqrt{\gamma_1}\,\|\mathbf{B}'\mathbf{q}_\mathrm{r}^*\|_2 \\ &= s^* + s_0 + \gamma_2\mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^* + \gamma_2\mathbf{I}_m \bullet \mathbf{Q}_0 + \sqrt{\gamma_1}\,\|\mathbf{q}_\mathrm{r}^*\|_2 \\ &= \underline{\Theta}(m_1, \mathbf{B}') + s_0 + \sum_{k=1}^K \frac{\gamma_2}{4(S + s_0)}\mathbf{I}_m \bullet \mathbf{M}_k, \end{aligned}$$

where the second equality holds because $\mathbf{I}_m \bullet \mathbf{B}'\mathbf{Q}_\mathrm{r}^*(\mathbf{B}')^\top = \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^*(\mathbf{B}')^\top\mathbf{B}' = \mathbf{I}_{m_1} \bullet \mathbf{Q}_\mathrm{r}^*$ and $(\mathbf{q}_\mathrm{r}^*)^\top(\mathbf{B}')^\top\mathbf{B}'\mathbf{q}_\mathrm{r}^* = (\mathbf{q}_\mathrm{r}^*)^\top\mathbf{q}_\mathrm{r}^*$. As this constructed solution is a feasible solution of

Problem (2.4), which is a minimization problem, we have $\Theta_{\mathrm{M}}(m) \le \Theta'_{\mathrm{M}}$ It follows that

$$\Theta_{\mathrm{M}}(m) - \Theta_{\mathrm{L}}(m_1) \le \Theta'_{\mathrm{M}} - \underline{\Theta}(m_1, \mathbf{B}') = s_0 + \sum_{k=1}^{K} \frac{\gamma_2}{4(S+s_0)} \mathbf{I}_m \bullet \mathbf{M}_k. \tag{A.63}$$

We further choose a value of s_0 to minimize the right-hand side (RHS) of (A.63). Note that (i) If $\sqrt{P} - S < 0$, then the RHS of (A.63) is minimized at P/S with $s_0 = 0$; (ii) If $\sqrt{P} - S \ge 0$, then the RHS of (A.63) is minimized at $2\sqrt{P} - S$ with $s_0 = \sqrt{P} - S$. Therefore, we conclude that the proposition holds. $\square$

A.7 Supplement to Section 2.7

A.7.1 Multiproduct Newsvendor Problem

By Proposition 1, Problem (2.39) has the same optimal value as the following SDP formulation:

$$\min_{\substack{\mathbf{x}, s, \boldsymbol{\lambda}_1, \\ \boldsymbol{\lambda}_2, \mathbf{q}, \mathbf{Q}}} \quad s + \gamma_2 \mathbf{I}_m \bullet \mathbf{Q} + \sqrt{\gamma_1} \left\| \mathbf{q} \right\|_2 \tag{A.64a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - (\mathbf{c} - \mathbf{v})^\top \mathbf{x} - \boldsymbol{\lambda}_1^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{A.64b}$$

$$\begin{bmatrix} s - (\mathbf{c} - \mathbf{g})^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v} - \mathbf{g})^\top \boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g})\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g})\right) & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{A.64c}$$

$$\mathbf{x} \in \mathbb{R}_+^m, \ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l, \ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l, \ \mathbf{q} \in \mathbb{R}^m, \ \mathbf{Q} \in \mathbb{R}^{m \times m}. \tag{A.64d}$$

By the first outer approximation (2.25), the following problem provides a lower bound for the optimal value of Problem (A.64):

$$\max_{\substack{\mathbf{B},t_1,\mathbf{p}_1,\mathbf{P}_1,\\ t_2,\mathbf{p}_2,\mathbf{P}_2}} \quad \left(t_2\boldsymbol{\mu}^\top + \mathbf{p}_2^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\right)(\mathbf{g}-\mathbf{v}) \tag{A.65a}$$

$$\text{s.t.} \quad 1 - t_1 - t_2 = 0,\ \sqrt{\gamma_1} - \|\mathbf{p}_1 + \mathbf{p}_2\|_2 \geq 0, \tag{A.65b}$$

$$t_1(\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_1^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top \leq 0, \tag{A.65c}$$

$$t_2(\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_2^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top \leq 0, \tag{A.65d}$$

$$\gamma_2\mathbf{I}_{m_1} - \mathbf{P}_1 - \mathbf{P}_2 \succeq 0,\ t_1(\mathbf{c}-\mathbf{v}) + t_2(\mathbf{c}-\mathbf{g}) \geq 0, \tag{A.65e}$$

$$\begin{bmatrix} t_1 & \mathbf{p}_1^\top \\ \mathbf{p}_1 & \mathbf{P}_1 \end{bmatrix} \succeq 0,\ \begin{bmatrix} t_2 & \mathbf{p}_2^\top \\ \mathbf{p}_2 & \mathbf{P}_2 \end{bmatrix} \succeq 0,\ \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}, \tag{A.65f}$$

$$\mathbf{B} \in \mathbb{R}^{m\times m_1},\ \mathbf{p}_1 \in \mathbb{R}^{m_1},\ \mathbf{p}_2 \in \mathbb{R}^{m_1},\ \mathbf{P}_1 \in \mathbb{R}^{m_1\times m_1},\ \mathbf{P}_2 \in \mathbb{R}^{m_1\times m_1}. \tag{A.65g}$$

By the inner approximation (2.26), the following problem provides an upper bound for the optimal value of Problem (A.64) and achieves the optimal value of Problem (A.64) when $m_1 \geq 2$:

$$\min_{\substack{\mathbf{B},\mathbf{x},s,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\\ \mathbf{q},\mathbf{Q}_\mathrm{r},\mathbf{u}_1,\mathbf{u}_2}} \quad s + \gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_\mathrm{r} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 \tag{A.66a}$$

$$\text{s.t.} \quad \begin{bmatrix} s - (\mathbf{c}-\mathbf{v})^\top\mathbf{x} - \boldsymbol{\lambda}_1^\top(\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{A.66b}$$

$$\begin{bmatrix} s - (\mathbf{c}-\mathbf{g})^\top\mathbf{x} - \boldsymbol{\lambda}_2^\top(\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v}-\mathbf{g})^\top\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0, \tag{A.66c}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\mathbf{A}^\top\boldsymbol{\lambda}_1 = \mathbf{B}\mathbf{u}_1, \tag{A.66d}$$

$$\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right) = \mathbf{B}\mathbf{u}_2, \tag{A.66e}$$

$$\mathbf{x} \in \mathbb{R}^m_+,\ \boldsymbol{\lambda}_1 \in \mathbb{R}^l_+,\ \boldsymbol{\lambda}_2 \in \mathbb{R}^l_+,\ \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}, \tag{A.66f}$$

$$\mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_\mathrm{r} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{B} \in \mathbb{R}^{m\times m_1},\ \mathbf{u}_1 \in \mathbb{R}^{m_1},\ \mathbf{u}_2 \in \mathbb{R}^{m_1}. \tag{A.66g}$$

By the second outer approximation (2.31), the following problem with $m_1 \leq 2$ provides another lower bound for the optimal value of Problem (A.64) and achieves the optimal value of Problem (A.64) when $m_1 = 2$:

$$
\begin{aligned}
\min_{\substack{\mathbf{B},\bar{\mathbf{B}},\mathbf{x},s,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\\ \mathbf{q},\mathbf{Q}_{\mathrm{r}},\mathbf{u}_1,\mathbf{u}_2,\mathbf{h}_1,\mathbf{h}_2}} \quad & s + \gamma_2 \mathbf{I}_{m_1} \bullet \mathbf{Q}_{\mathrm{r}} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 && \text{(A.67a)}\\
\text{s.t.} \quad & \begin{bmatrix} s - (\mathbf{c}-\mathbf{v})^\top \mathbf{x} - \boldsymbol{\lambda}_1^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) & \frac{1}{2}\mathbf{u}_1^\top \\ \frac{1}{2}\mathbf{u}_1 & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, && \text{(A.67b)}\\
& \begin{bmatrix} s - (\mathbf{c}-\mathbf{g})^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top (\mathbf{b} - \mathbf{A}\boldsymbol{\mu}) + (\mathbf{v}-\mathbf{g})^\top \boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_2^\top \\ \frac{1}{2}\mathbf{u}_2 & \mathbf{Q}_{\mathrm{r}} \end{bmatrix} \succeq 0, && \text{(A.67c)}\\
& \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \mathbf{A}^\top \boldsymbol{\lambda}_1 = \mathbf{B}\mathbf{u}_1 + \bar{\mathbf{B}}\mathbf{h}_1, && \text{(A.67d)}\\
& \mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top \left(\mathbf{A}^\top \boldsymbol{\lambda}_2 + \mathbf{v} - \mathbf{g}\right) = \mathbf{B}\mathbf{u}_2 + \bar{\mathbf{B}}\mathbf{h}_2, && \text{(A.67e)}\\
& \mathbf{x} \in \mathbb{R}_+^m,\ \boldsymbol{\lambda}_1 \in \mathbb{R}_+^l,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^l,\ [\mathbf{B},\bar{\mathbf{B}}]^\top [\mathbf{B},\bar{\mathbf{B}}] = \mathbf{I}_K, && \text{(A.67f)}\\
& \mathbf{q} \in \mathbb{R}^m,\ \mathbf{Q}_{\mathrm{r}} \in \mathbb{R}^{m_1\times m_1},\ \mathbf{B} \in \mathbb{R}^{m\times m_1},\ \bar{\mathbf{B}} \in \mathbb{R}^{m\times (K-m_1)}, && \text{(A.67g)}\\
& \mathbf{u}_1 \in \mathbb{R}^{m_1},\ \mathbf{u}_2 \in \mathbb{R}^{m_1},\ \mathbf{h}_1 \in \mathbb{R}^{K-m_1},\ \mathbf{h}_2 \in \mathbb{R}^{K-m_1}. && \text{(A.67h)}
\end{aligned}
$$

A.7.2 Production-Transportation Problem

By the reformulation results of Section 4.1 of Bertsimas et al. (2010) and Appendix C of Cheramin et al. (2022), we have the following SDP reformulation for Problem (2.41):

$$
\begin{aligned}
\min_{\substack{\mathbf{x},\mathbf{z}_k(\forall k\in[K]),\\ s,\boldsymbol{\lambda}_k(\forall k\in[K]),\mathbf{q},\mathbf{Q}}} \quad & s + \gamma_2 \mathbf{I}_{GH} \bullet \mathbf{Q} + \sqrt{\gamma_1}\,\|\mathbf{q}\|_2 && \text{(A.68a)}\\
\text{s.t.} \quad & \begin{bmatrix} s - \mathbf{c}^\top \mathbf{x} - \beta_k - \boldsymbol{\lambda}_k^\top \mathbf{b} - \alpha_k \mathbf{z}_k^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_k^\top \mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k)\right)^\top \\ \frac{1}{2}\left(\mathbf{q} + \left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top (\mathbf{A}^\top \boldsymbol{\lambda}_k - \alpha_k \mathbf{z}_k)\right) & \mathbf{Q} \end{bmatrix} \succeq 0,\\
& \hspace{20em} \forall k \in [K], && \text{(A.68b)}\\
& \boldsymbol{\lambda}_k \in \mathbb{R}_+^l,\ \forall k \in [K], \mathbf{q} \in \mathbb{R}^{GH}, \mathbf{Q} \in \mathbb{R}^{GH\times GH}, && \text{(A.68c)}\\
& \mathbf{0} \le \mathbf{x} \le \mathbf{1}, && \text{(A.68d)}\\
& \sum_{i=1}^{G} z_{ijk} = d_j,\ \forall j \in [H],\ k \in [K], && \text{(A.68e)}
\end{aligned}
$$

$$\sum_{j=1}^{H} z_{ijk} = x_i, \ \forall i \in [G], \ k \in [K], \tag{A.68f}$$

$$z_{ijk} \geq 0, \ \forall i \in [G], \ j \in [H], \ k \in [K], \tag{A.68g}$$

where $\mathbf{z}_k$ is a vector whose $((i-1)G+j)$-th element is z_{ijk}. Note that, by Theorem 1 in Cheramin et al. (2022), the following problem can be reformulated as Problem (A.68):

$$\min_{\mathbf{x},\mathbf{z}_k,\forall k\in[K]} \max_{\mathbb{P}_{\mathrm{I}}\in\mathcal{D}_{\mathrm{M}}} \quad \mathbb{E}_{\mathbb{P}}\left[\max_{k\in[K]}\left\{\mathbf{c}^\top\mathbf{x} + \alpha_k\mathbf{z}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}_{\mathrm{I}} + \boldsymbol{\mu}\right) + \beta_k\right\}\right] \tag{A.69a}$$

$$\text{s.t.} \quad (\text{A.68d}) - (\text{A.68g}). \tag{A.69b}$$

By the first outer approximation (2.25), the following problem provides a lower bound for the optimal value of Problem (A.68):

$$\max_{\substack{t_k,\mathbf{p}_k,\mathbf{P}_k,\forall k\in[K],\\ \mathbf{w}_k,\mathbf{u}_k,\forall k\in[K],\mathbf{v},\mathbf{B}}} \quad \sum_{k=1}^{K} t_k\beta_k - \sum_{i=i}^{G} v_i + \sum_{k=1}^{K}\sum_{j=1}^{H} w_{jk}d_j \tag{A.70a}$$

$$\text{s.t.} \quad 1 - \sum_{k=1}^{K} t_k = 0, \ \sqrt{\gamma_1} - \left\|\sum_{k=1}^{K}\mathbf{p}_k\right\|_2 \geq 0, \tag{A.70b}$$

$$t_k(\mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top + \mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top\mathbf{A}^\top \leq 0, \ \forall k \in [K], \tag{A.70c}$$

$$\gamma_2\mathbf{I}_{m_1} - \sum_{k=1}^{K}\mathbf{P}_k \succeq 0, \ \sum_{k=1}^{K}(t_k\mathbf{c} + \mathbf{u}_k) + \mathbf{v} \geq 0, \tag{A.70d}$$

$$\begin{aligned}&\alpha_k t_k\boldsymbol{\mu}^\top + \alpha_k\mathbf{p}_k^\top\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{B}\right)^\top - (\underbrace{\mathbf{w}_k^\top,\ldots,\mathbf{w}_k^\top}_{\text{repeat G times}})\\ &\quad -(\underbrace{u_{1k},\ldots,u_{1k}}_{\text{repeat H times}},\ldots,u_{Gk},\ldots,u_{Gk}) \geq 0, \forall k \in [K],\end{aligned} \tag{A.70e}$$

$$\begin{bmatrix} t_k & \mathbf{p}_k^\top \\ \mathbf{p}_k & \mathbf{P}_k \end{bmatrix} \succeq 0, \ \forall k \in [K], \ \mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}, \ \mathbf{v} \in \mathbb{R}_+^G, \ \mathbf{B} \in \mathbb{R}^{GH\times m_1}, \tag{A.70f}$$

$$\mathbf{p}_k \in \mathbb{R}^{m_1}, \ \mathbf{P}_k \in \mathbb{R}^{m_1\times m_1}, \ \mathbf{w}_k \in \mathbb{R}^H, \ \mathbf{u}_k \in \mathbf{R}^G, \ \forall k \in [K]. \tag{A.70g}$$

By the inner approximation (2.26), the following problem provides an upper bound for the optimal value of Problem (A.68) and achieves the optimal value of Problem (A.68) when $m_1 \geq K$:

$$\min_{\substack{\mathbf{B},\mathbf{x},\mathbf{z}_k(\forall k\in[K]),\\ \mathbf{u}_k,\boldsymbol{\lambda}_k,\forall k\in[K],s,\mathbf{q},\mathbf{Q}_\mathrm{r}}} \quad s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_\mathrm{r}+\sqrt{\gamma_1}\left\|\mathbf{q}\right\|_2 \tag{A.71a}$$

$$\text{s.t.} \quad \begin{bmatrix} s-\mathbf{c}^\top\mathbf{x}-\beta_k-\boldsymbol{\lambda}_k^\top\mathbf{b}-\alpha_k\mathbf{z}_k^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0,\ \forall k\in[K], \tag{A.71b}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k-\alpha_k\mathbf{z}_k)\right)=\mathbf{B}\mathbf{u}_k,\ \forall k\in[K], \tag{A.71c}$$

$$\boldsymbol{\lambda}_k\in\mathbb{R}_+^l,\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \forall k\in[K],\mathbf{q}\in\mathbb{R}^{GH},\mathbf{Q}_\mathrm{r}\in\mathbb{R}^{m_1\times m_1}, \tag{A.71d}$$

$$(\text{A.68d})-(\text{A.68g}),\ \mathbf{B}^\top\mathbf{B}=\mathbf{I}_{m_1},\ \mathbf{B}\in\mathbb{R}^{GH\times m_1}. \tag{A.71e}$$

By the second outer approximation (2.31), the following problem with $m_1 \leq K$ provides another lower bound for the optimal value of Problem (A.68) and achieves the optimal value of Problem (A.68) when $m_1 = K$:

$$\min_{\substack{\mathbf{B},\bar{\mathbf{B}},\mathbf{x},\mathbf{z}_k(\forall k\in[K]),\\ \mathbf{u}_k,\mathbf{h}_k,\boldsymbol{\lambda}_k,\forall k\in[K],\\ s,\mathbf{q},\mathbf{Q}_\mathrm{r}}} \quad s+\gamma_2\mathbf{I}_{m_1}\bullet\mathbf{Q}_\mathrm{r}+\sqrt{\gamma_1}\left\|\mathbf{q}\right\|_2 \tag{A.72a}$$

$$\text{s.t.} \begin{bmatrix} s-\mathbf{c}^\top\mathbf{x}-\beta_k-\boldsymbol{\lambda}_k^\top\mathbf{b}-\alpha_k\mathbf{z}_k^\top\boldsymbol{\mu}+\boldsymbol{\lambda}_k^\top\mathbf{A}\boldsymbol{\mu} & \frac{1}{2}\mathbf{u}_k^\top \\ \frac{1}{2}\mathbf{u}_k & \mathbf{Q}_\mathrm{r} \end{bmatrix} \succeq 0,\ \forall k\in[K], \tag{A.72b}$$

$$\mathbf{q}+\left(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^\top\left(\mathbf{A}^\top\boldsymbol{\lambda}_k-\alpha_k\mathbf{z}_k)\right)=\mathbf{B}\mathbf{u}_k+\bar{\mathbf{B}}\mathbf{h}_k,\ \forall k\in[K], \tag{A.72c}$$

$$\boldsymbol{\lambda}_k\in\mathbb{R}_+^l,\ \mathbf{u}_k\in\mathbb{R}^{m_1},\ \mathbf{h}_k\in\mathbb{R}^{K-m_1},\ \forall k\in[K],\mathbf{q}\in\mathbb{R}^{GH},\mathbf{Q}_\mathrm{r}\in\mathbb{R}^{m_1\times m_1}, \tag{A.72d}$$

$$[\mathbf{B},\bar{\mathbf{B}}]^\top[\mathbf{B},\bar{\mathbf{B}}]=\mathbf{I}_{m_1},\ \mathbf{B}\in\mathbb{R}^{GH\times m_1},\ \bar{\mathbf{B}}\in\mathbb{R}^{GH\times(K-m_1)}, \tag{A.72e}$$

$$(\text{A.68d})-(\text{A.68g}). \tag{A.72f}$$

A.7.3 Proof of Proposition 7

Because $\mathbf{B}^\top\mathbf{B} = \mathbf{I}_{m_1}$, we have $\mathbf{B}^\top\mathbf{B} \preceq \mathbf{I}_{m_1}$, which implies $\mathbf{B}\mathbf{B}^\top \preceq \mathbf{I}_m$ by Lemma 1. It follows that $\mathbf{r}^\top\mathbf{B}\mathbf{B}^\top\mathbf{r} \leq \mathbf{r}^\top\mathbf{r}$. Meanwhile, we have

$$\mathbf{r}^\top\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{r} = \begin{bmatrix} \frac{\mathbf{r}^\top\mathbf{r}}{\|\mathbf{r}\|_2} & \mathbf{0}_{1\times(m_1-1)} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{r}^\top\mathbf{r}}{\|\mathbf{r}\|_2} \\ \mathbf{0}_{(m_1-1)\times 1} \end{bmatrix} = \mathbf{r}^\top\mathbf{r},$$

indicating that $\mathbf{B}^* = \begin{bmatrix} \mathbf{r}/\|\mathbf{r}\|_2, \mathbf{0}_{m\times(m_1-1)} \end{bmatrix}$ is an optimal solution of Problem (2.44). $\square$

A.7.4 Sensitivity Analyses

Table A.3: Sensitivity Analyses on the Production-Transportation Problem with $K = 5$

		Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 3$	ODR-LB	Gap1 (%)	0.37	0.71	0.67	0.47	-	-	-
		Time (secs)	2.86	2.81	4.87	4.32	15.16	27.51	57.45
		Interval Gap (%)	0.69	0.79	0.73	0.51	0.49	0.51	0.57
		Theoretical Gap (%)	1.27	0.62	2.20	1.36	-	-	-
	ODR-RLB	Gap1 (%)	0.07	0.17	0.02	0.02	-	-	-
		Time (secs)	13.53	14.19	43.93	57.79	205.99	442.26	1092.43
		Interval Gap (%)	0.39	0.25	0.08	0.06	0.03	0.02	0.02
		Theoretical Gap (%)	1.08	3.19	1.96	0.96	-	-	-
	ODR-UB	Gap2 (%)	0.32	0.08	0.06	0.04	-	-	-
		Time (secs)	6.03	5.60	22.31	20.53	122.04	332.10	660.08
		Theoretical Gap (%)	1.28	1.03	3.20	2.05	-	-	-
$m_1 = 5$	ODR-LB	Gap1 (%)	0.14	0.34	0.22	0.29	-	-	-
		Time (secs)	4.03	3.63	6.01	4.82	12.41	25.03	57.79
		Interval Gap (%)	0.15	0.34	0.43	0.29	0.52	0.55	0.51
		Theoretical Gap (%)	1.24	1.07	4.91	0.77	2.71	1.25	1.09
	ODR-RLB	Gap1 (%)	0.01	0.02	0.01	0.00	-	-	-
		Time (secs)	5.42	5.36	22.40	20.86	123.64	330.29	665.43
		Interval Gap (%)	0.02	0.02	0.01	0.01	0.00	0.00	0.00
		Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
	ODR-UB	Gap2 (%)	0.01	0.01	0.00	0.00	-	-	-
		Time (secs)	5.48	5.32	22.41	20.86	123.48	329.95	665.35
		Theoretical Gap (%)	1.13	1.17	1.80	0.81	1.22	1.92	1.32
$m_1 = 7$	ODR-LB	Gap1 (%)	0.16	0.21	0.53	0.23	-	-	-
		Time (secs)	3.56	4.29	5.39	5.70	19.27	22.89	58.38
		Interval Gap (%)	0.17	0.22	0.53	0.24	0.31	0.56	0.50
		Theoretical Gap (%)	1.22	0.88	1.67	0.79	-	-	-
	ODR-UB	Gap2 (%)	0.02	0.01	0.00	0.00	-	-	-
		Time (secs)	5.62	5.48	22.42	21.24	121.61	330.94	660.85
		Theoretical Gap (%)	1.15	1.04	1.28	0.59	-	-	-

Table A.4: Sensitivity Analyses on the Production-Transportation Problem with $K = 10$

		Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 8$	ODR-LB	Gap1 (%)	0.17	0.15	0.16	0.28	-	-	-
		Time (secs)	6.26	5.75	9.80	8.87	29.94	54.71	133.58
		Interval Gap (%)	0.17	0.15	0.16	0.28	0.23	0.22	0.27
		Theoretical Gap (%)	4.32	4.53	6.22	6.09	-	-	-
	ODR-RLB	Gap1 (%)	0.01	0.00	0.01	0.00	-	-	-
		Time (secs)	13.03	12.91	53.84	35.92	205.07	593.38	1334.02
		Interval Gap (%)	0.01	0.00	0.01	0.01	0.00	0.00	0.00
		Theoretical Gap (%)	3.53	3.02	4.24	4.86	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	10.31	10.23	39.26	31.59	119.58	339.84	755.18
		Theoretical Gap (%)	2.21	2.40	3.98	3.04	-	-	-
$m_1 = 10$	ODR-LB	Gap1 (%)	0.12	0.19	0.13	0.25	-	-	-
		Time (secs)	6.65	6.41	11.11	10.31	31.42	61.41	131.10
		Interval Gap (%)	0.12	0.19	0.13	0.25	0.22	0.19	0.25
		Theoretical Gap (%)	1.78	1.97	2.96	1.19	2.10	1.49	1.22
	ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.22	10.65	40.34	32.95	122.54	342.80	748.11
		Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.14	10.69	40.28	32.92	122.48	344.10	747.90
		Theoretical Gap (%)	1.86	2.27	1.74	1.22	2.10	1.39	2.30
$m_1 = 12$	ODR-LB	Gap1 (%)	0.11	0.14	0.15	0.23	-	-	-
		Time (secs)	8.89	7.90	11.78	13.44	32.16	63.96	122.91
		Interval Gap (%)	0.11	0.14	0.15	0.23	0.19	0.19	0.27
		Theoretical Gap (%)	3.64	5.01	7.41	7.85	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	11.71	11.55	38.59	29.68	123.40	346.68	761.21
		Theoretical Gap (%)	0.81	1.32	2.20	2.75	-	-	-

Table A.5: Sensitivity Analyses on the Production-Transportation Problem with $K = 15$

		Size $((G, H))$	(4,25)	(5,20)	(5,40)	(8,25)	(10,40)	(20,30)	(20,40)
$m_1 = 13$	ODR-LB	Gap1 (%)	0.08	0.21	0.15	0.18	-	-	-
		Time (secs)	11.12	11.10	18.00	24.05	31.32	71.04	142.88
		Interval Gap (%)	0.08	0.21	0.15	0.18	0.25	0.19	0.23
		Theoretical Gap (%)	4.56	7.01	5.79	6.95	-	-	-
	ODR-RLB	Gap1 (%)	0.00	0.03	0.00	0.00	-	-	-
		Time (secs)	26.19	25.91	97.35	68.99	213.44	641.74	1451.80
		Interval Gap (%)	0.00	0.03	0.00	0.00	0.00	0.01	0.00
		Theoretical Gap (%)	6.09	5.83	5.88	6.12	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	19.46	19.15	71.71	50.62	128.92	367.33	833.81
		Theoretical Gap (%)	2.44	3.00	2.52	3.07	-	-	-
$m_1 = 15$	ODR-LB	Gap1 (%)	0.05	0.10	0.12	0.15	-	-	-
		Time (secs)	13.92	16.56	22.11	26.16	43.02	80.51	168.90
		Interval Gap (%)	0.05	0.10	0.12	0.15	0.11	0.13	0.09
		Theoretical Gap (%)	4.69	5.22	5.57	6.38	3.29	4.41	3.95
	ODR-RLB	Gap1 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	22.60	21.45	77.18	63.99	241.03	689.24	1550.21
		Interval Gap (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	22.61	21.44	76.99	64.08	149.21	401.63	878.68
		Theoretical Gap (%)	1.38	2.61	1.79	1.73	2.14	1.94	2.25
$m_1 = 17$	ODR-LB	Gap1 (%)	0.10	0.09	0.13	0.16	-	-	-
		Time (secs)	18.10	21.04	28.81	32.61	52.71	84.10	188.21
		Interval Gap (%)	0.10	0.09	0.13	0.16	0.10	0.08	0.13
		Theoretical Gap (%)	4.75	4.13	5.50	6.16	-	-	-
	ODR-UB	Gap2 (%)	0.00	0.00	0.00	0.00	-	-	-
		Time (secs)	26.83	27.84	84.70	63.17	161.79	447.03	926.33
		Theoretical Gap (%)	2.20	3.61	2.58	2.20	-	-	-

Appendix B

Supplements for Chapter 3

B.1 Notation and Abbreviations

Table B.1: Summary of Notation and Abbreviations

Notation/Abbreviation	Description
Decision Variables:	
$\mathbf{x}$	The decision variable
$\mathbf{y}_i$, $\mathbf{z}_i$, $\mathbf{t}_i$	The auxiliary decision variables
$\mathbf{u}$, $\mathbf{w}$, $\mathbf{s}$, $\mathbf{v}$	The extra variables used in approximations
Parameters:	
$\mathbf{c}_0$	The coefficient of the objective function
$\mathbf{Q}_i$, $\mathbf{c}_i$, $\mathbf{b}_i$	The parameters in constraints
$\mathbf{A}_i$, $\mathbf{B}_i$	The parameters obtained by decomposing $\mathbf{Q}_i$
U, U_y, U_z	The upper bounds of decision variables
ε	The quality parameter of approximations
$\mathbf{P}$, $\mathbf{H}$, $\mathbf{p}$	The parameters in the first polyhedral approximation

Optimal Values:

Θ^*	The optimal value of the original non-convex QCP (3.1)
Θ_ε	The optimal value of the LPCC/MILP approximations
$\underline{\Theta}_\varepsilon$	The optimal value of the "ε-relaxation" of the original non-convex QCP (3.1)

Abbreviations:

LP	Linear Program
MILP	Mixed-Integer Linear Program
QCP	Quadratically Constrained Program
SOC	Second-Order Cone
SOCP	Second-Order Cone Program
SDP	Semidefinite Program
MISOCP	Mixed-Integer Second-Order Cone Program
LPCC	Linear Program with Complementarity Constraints
RLT	Reformulation-Linearization Technique
ACOPF	AC Optimal Power Flow
JDE	Joint Decision and Estimation
TTRS	Two-Trust-Region Subproblem
P1	The first polyhedral approximation (Section 3.2.2)
P2	The second polyhedral approximation (Section 3.3.1)
S-P1	The simplified version of the first polyhedral approximation (Section 3.4.1)
A1	The first LPCC/MILP approximation (Problem (3.11)/(3.31))

B.2 Boundedness of $\mathbf{y}_i$ and $\mathbf{z}_i$ for any $i \in [m]$

For each $i \in [m]$, we have

$$\|\mathbf{y}_i\| = \sqrt{\mathbf{x}^\top \mathbf{A}_i \mathbf{A}_i^\top \mathbf{x} + \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} - 1}{2}\right)^2} \leq \sqrt{\lambda_{\max}(\mathbf{Q}_i) U^2 + \left(\frac{1}{2}|b_i - 1| + \|\mathbf{c}_i\| U\right)^2} := U_y^i,$$
$$\|\mathbf{z}_i\| = \sqrt{\mathbf{x}^\top \mathbf{B}_i \mathbf{B}_i^\top \mathbf{x} + \left(\frac{b_i - 2\mathbf{c}_i^\top \mathbf{x} + 1}{2}\right)^2} \leq \sqrt{-\lambda_{\min}(\mathbf{Q}_i) U^2 + \left(\frac{1}{2}|b_i + 1| + \|\mathbf{c}_i\| U\right)^2} := U_z^i,$$

where $\lambda_{\max}(\mathbf{Q}_i)$ and $\lambda_{\min}(\mathbf{Q}_i)$ denote the largest and smallest eigenvalues of $\mathbf{Q}_i$, respectively. Therefore, we can set $U_y = \max\{U_y^i, \forall i \in [m]\}$ and $U_z = \max\{U_z^i, \forall i \in [m]\}$. Note that the upper bounds on $\|\mathbf{y}_i\|$ and $\|\mathbf{z}_i\|$ can be refined by solving trust-region problems. Specifically, an upper bound of $\mathbf{y}_i^\top \mathbf{y}_i$ can be computed by solving

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^\top \left(\mathbf{A}_i \mathbf{A}_i^\top + \mathbf{c}_i \mathbf{c}_i^\top\right) \mathbf{x} - \frac{b_i - 1}{2} \mathbf{c}_i^\top \mathbf{x} + \left(\frac{b_i - 1}{2}\right)^2 \\ \text{s.t.} \quad & \mathbf{x}^\top \mathbf{x} \leq U^2. \end{aligned}$$

Similarly, an upper bound of $\mathbf{z}_i^\top \mathbf{z}_i$ can be computed by solving

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^\top \left(\mathbf{B}_i \mathbf{B}_i^\top + \mathbf{c}_i \mathbf{c}_i^\top\right) \mathbf{x} - \frac{b_i + 1}{2} \mathbf{c}_i^\top \mathbf{x} + \left(\frac{b_i + 1}{2}\right)^2 \\ \text{s.t.} \quad & \mathbf{x}^\top \mathbf{x} \leq U^2. \end{aligned}$$

B.3 An Alternative Polyhedral Approximation of $\|\mathbf{y}\| \leq t$

First, we consider

$$L^2 = \left\{ (y_1, y_2, y_3) \;\middle|\; \sqrt{y_1^2 + y_2^2} \leq y_3 \right\}.$$

With a given integer number $\nu \geq 1$, we provide an alternative polyhedral approximation of the set L^2 using the following constraints:

$$\xi_0 \geq |y_1|, \tag{B.1a}$$

$$\eta_0 \geq |y_2|, \tag{B.1b}$$

$$\cos\left(\frac{\tau\pi}{2^\nu}\right)\xi_0 + \sin\left(\frac{\tau\pi}{2^\nu}\right)\eta_0 \leq y_3,\ \forall\, \tau \in \{0\} \cup [2^{\nu-1}]. \tag{B.1c}$$

This polyhedral approximation of the set L^2 can be described as

$$\Pi^2(\nu) := \{(y_1, y_2, y_3) \mid \text{(B.1) is satisfied for some } \xi_0 \text{ and } \eta_0\}.$$

Next, the same as (3.4), we extend the set L^2 to the following set L^k with $k \geq 2$:

$$L^k = \left\{ (\mathbf{y}, t) \in \mathbb{R}^k \times \mathbb{R} \;\middle|\; \begin{array}{l} \sqrt{\left(y_{2j-1}^{\ell-1}\right)^2 + \left(y_{2j}^{\ell-1}\right)^2} \leq y_j^\ell,\ \forall\, j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta] \\ \text{for some } y_{k+1}^0, \ldots, y_{2^\theta}^0, \text{ and } y_j^\ell,\ \forall j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta-1] \end{array} \right\}.$$

Note that $y_{k+1}^0, \ldots, y_{2^\theta}^0$ and y_j^l ($\forall j \in [2^{\theta-\ell}], \ell \in [\theta-1]$) are extra variables and not included in the vector $\mathbf{y}$. Using $\Pi^2(\nu)$ to approximate L^2, we obtain the following polyhedral approximation of L^k:

$$\begin{aligned} &\Pi^k(\nu_1, \ldots, \nu_\theta) \\ &:= \left\{ (\mathbf{y}, t) \in \mathbb{R}^k \times \mathbb{R} \;\middle|\; \begin{array}{l} \left(y_{2j-1}^{\ell-1}, y_{2j}^{\ell-1}, y_j^\ell\right) \in \Pi^2(\nu_\ell),\ \forall\, j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta] \\ \text{for some } y_{k+1}^0, \ldots, y_{2^\theta}^0, \text{ and } y_j^\ell,\ \forall j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta-1] \end{array} \right\}, \end{aligned}$$

$$= \left\{ (\mathbf{y}, t) \;\middle|\; \begin{array}{l} \xi_j \geq |y^0_{2j-1}|, \eta_j \geq |y^0_{2j}|, \xi_j \boldsymbol{\alpha}(\nu_1) + \eta_j \boldsymbol{\beta}(\nu_1) \leq y^1_j \mathbf{e},\ \forall\, j \in \left[2^{\theta-1}\right], \\ y^{\ell-1}_{2j-1} \boldsymbol{\alpha}(\nu_\ell) + y^{\ell-1}_{2j} \boldsymbol{\beta}(\nu_\ell) \leq y^\ell_j \mathbf{e},\ \forall\, j \in \left[2^{\theta-\ell}\right],\ \ell \in \{2, \ldots, \theta\} \\ \text{for some } \xi_j, \eta_j,\ \forall j \in [2^{\theta-1}],\ y^0_{k+1}, \ldots, y^0_{2^\theta},\ \text{and } y^\ell_j,\ \forall j \in \left[2^{\theta-\ell}\right],\ \ell \in [\theta-1] \end{array} \right\},$$

where $\boldsymbol{\alpha}(\nu_\ell) := (\cos(\tau\pi/2^{\nu_\ell}))_{\tau=0}^{2^{\nu_\ell-1}} \in \mathbb{R}_+^{2^{\nu_\ell-1}+1}$, $\boldsymbol{\beta}(\nu_\ell) := (\sin(\tau\pi/2^{\nu_\ell}))_{\tau=0}^{2^{\nu_\ell-1}} \in \mathbb{R}_+^{2^{\nu_\ell-1}+1}$, and ν_ℓ ($\forall \ell \in \theta$) are selected to control the approximation quality ε_ℓ in level ℓ.

B.4 Proof of Theorem 8

By Lemma 4, when the dimension of $\mathbf{z}$ is 2^1, the feasible region of Problem (3.9) is bounded.

When the dimension of $\mathbf{z}$ is 2^θ, Problem (3.9) becomes:

$$\min \quad \lambda^6_{\theta,1} \tag{B.2a}$$

$$\text{s.t.} \quad \lambda^1_{1,i} - \lambda^2_{1,i} = z_{2i-1},\ \forall\, i \in [2^{\theta-1}], \tag{B.2b}$$

$$\lambda^3_{1,i} - \lambda^4_{1,i} = z_{2i},\ \forall\, i \in [2^{\theta-1}], \tag{B.2c}$$

$$\lambda^1_{\ell,i} - \lambda^2_{\ell,i} - \lambda^6_{\ell-1,2i-1} = 0,\ \forall\, \ell \in [\theta] \setminus \{1\}, i \in [2^{\theta-\ell}], \tag{B.2d}$$

$$\lambda^3_{\ell,i} - \lambda^4_{\ell,i} - \lambda^6_{\ell-1,2i} = 0,\ \forall\, \ell \in [\theta] \setminus \{1\}, i \in [2^{\theta-\ell}], \tag{B.2e}$$

$$-\lambda^1_{\ell,i} - \lambda^2_{\ell,i} + w^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) - k^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) + u^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) = 0, \tag{B.2f}$$

$$\forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \tag{B.2g}$$

$$-\lambda^3_{\ell,i} - \lambda^4_{\ell,i} + w^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) + k^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) - u^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) = 0, \tag{B.2h}$$

$$\forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \tag{B.2i}$$

$$-w^j_{\ell,i} + w^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) - k^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) + u^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) = 0,$$

$$\forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], j \in [\nu - 1], \tag{B.2j}$$

$$-k^j_{\ell,i} - u^j_{\ell,i} + w^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) + k^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) - u^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) = 0,$$

$$\forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], j \in [\nu - 1], \tag{B.2k}$$

$$-w^{\nu}_{\ell,i} - \tan\left(\frac{\pi}{2^{\nu+1}}\right)\lambda^5_{\ell,i} + \lambda^6_{\ell,i} = 0,\ \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \tag{B.2l}$$

$$-k^{\nu}_{\ell,i} - u^{\nu}_{\ell,i} + \lambda^5_{\ell,i} = 0,\ \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \tag{B.2m}$$

$$\lambda^1_{\ell,i}, \lambda^2_{\ell,i}, \lambda^3_{\ell,i}, \lambda^4_{\ell,i}, \lambda^5_{\ell,i}, \lambda^6_{\ell,i} \geq 0,\ \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \tag{B.2n}$$

$$k^j_{\ell,i}, u^j_{\ell,i} \geq 0,\ \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], j \in [\nu], \tag{B.2o}$$

$$\lambda^6_{\theta,1} \leq M_\theta. \tag{B.2p}$$

Now we assume that the feasible region of Problem (B.2) is bounded for any $M_\theta > 0$. We only need to prove that, when the dimension of $\mathbf{z}$ is $2^{\theta+1}$, for any $M_{\theta+1} > 0$, the feasible region of Problem (3.9) is bounded. Specifically, we should prove the feasible region of the following problem is bounded:

$$\min \quad \lambda^6_{\theta+1,1} \tag{B.3a}$$

$$\text{s.t.} \quad \text{(B.4)},\ \forall i \in [2^{\theta-1}], \text{(B.5)},\ \forall i \in [2^{\theta-\ell}], \tag{B.3b}$$

$$\text{(B.4)},\ \forall i \in \{2^{\theta-1}+1, \ldots, 2^{\theta}\}, \text{(B.5)},\ \forall i \in \{2^{\theta-\ell}+1, \ldots, 2^{\theta+1-\ell}\}, \tag{B.3c}$$

$$\lambda^1_{\theta+1,1} - \lambda^2_{\theta+1,1} - \lambda^6_{\theta,1} = 0, \tag{B.3d}$$

$$\lambda^3_{\theta+1,1} - \lambda^4_{\theta+1,1} - \lambda^6_{\theta,2} = 0, \tag{B.3e}$$

$$-\lambda^1_{\theta+1,1} - \lambda^2_{\theta+1,1} + w^1_{\theta+1,1}\cos\left(\frac{\pi}{4}\right) - k^1_{\theta+1,1}\sin\left(\frac{\pi}{4}\right) + u^1_{\theta+1,1}\sin\left(\frac{\pi}{4}\right) = 0, \tag{B.3f}$$

$$-\lambda^3_{\theta+1,1} - \lambda^4_{\theta+1,1} + w^1_{\theta+1,1}\sin\left(\frac{\pi}{4}\right) + k^1_{\theta+1,1}\cos\left(\frac{\pi}{4}\right) - u^1_{\theta+1,1}\cos\left(\frac{\pi}{4}\right) = 0, \tag{B.3g}$$

$$-w^j_{\theta+1,1} + w^{j+1}_{\theta+1,1}\cos\left(\frac{\pi}{2^{j+2}}\right) - k^{j+1}_{\theta+1,1}\sin\left(\frac{\pi}{2^{j+2}}\right) + u^{j+1}_{\theta+1,1}\sin\left(\frac{\pi}{2^{j+2}}\right) = 0,\ \forall j \in [\nu-1], \tag{B.3h}$$

$$-k^j_{\theta+1,1} - u^j_{\theta+1,1} + w^{j+1}_{\theta+1,1}\sin\left(\frac{\pi}{2^{j+2}}\right) + k^{j+1}_{\theta+1,1}\cos\left(\frac{\pi}{2^{j+2}}\right) - u^{j+1}_{\theta+1,1}\cos\left(\frac{\pi}{2^{j+2}}\right) = 0,\ \forall j \in [\nu-1], \tag{B.3i}$$

$$-w^{\nu}_{\theta+1,1} - \tan\left(\frac{\pi}{2^{\nu+1}}\right)\lambda^5_{\theta+1,1} + \lambda^6_{\theta+1,1} = 0, \tag{B.3j}$$

$$-k^{\nu}_{\theta+1,1} - u^{\nu}_{\theta+1,1} + \lambda^5_{\theta+1,1} = 0, \tag{B.3k}$$

$$\lambda^1_{\theta+1,1}, \lambda^2_{\theta+1,1}, \lambda^3_{\theta+1,1}, \lambda^4_{\theta+1,1}, \lambda^5_{\theta+1,1}, \lambda^6_{\theta+1,1} \geq 0, \tag{B.3l}$$

$$k^j_{\theta+1,1}, w^j_{\theta+1,1} \geq 0,\ \forall j \in [\nu], \tag{B.3m}$$

$$\lambda^6_{\theta+1,1} \leq M_{\theta+1}, \tag{B.3n}$$

where (B.4) and (B.5) are denoted as follows:

$$\lambda^1_{1,i} - \lambda^2_{1,i} = z_{2i-1}, \tag{B.4a}$$

$$\lambda^3_{1,i} - \lambda^4_{1,i} = z_{2i}. \tag{B.4b}$$

$$\lambda^1_{\ell,i} - \lambda^2_{\ell,i} - \lambda^6_{\ell-1,2i-1} = 0,\ \forall\, \ell \in [\theta] \setminus \{1\}, \tag{B.5a}$$

$$\lambda^3_{\ell,i} - \lambda^4_{\ell,i} - \lambda^6_{\ell-1,2i} = 0,\ \forall\, \ell \in [\theta] \setminus \{1\}, \tag{B.5b}$$

$$-\lambda^1_{\ell,i} - \lambda^2_{\ell,i} + w^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) - k^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) + u^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) = 0,\ \forall\, \ell \in [\theta], \tag{B.5c}$$

$$-\lambda^3_{\ell,i} - \lambda^4_{\ell,i} + w^1_{\ell,i} \sin\left(\frac{\pi}{4}\right) + k^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) - u^1_{\ell,i} \cos\left(\frac{\pi}{4}\right) = 0,\ \forall\, \ell \in [\theta], \tag{B.5d}$$

$$-w^j_{\ell,i} + w^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) - k^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) + u^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) = 0, \quad \forall \ell \in [\theta], j \in [\nu-1], \tag{B.5e}$$

$$-k^j_{\ell,i} - u^j_{\ell,i} + w^{j+1}_{\ell,i} \sin\left(\frac{\pi}{2^{j+2}}\right) + k^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) - u^{j+1}_{\ell,i} \cos\left(\frac{\pi}{2^{j+2}}\right) = 0, \quad \forall \ell \in [\theta], j \in [\nu-1], \tag{B.5f}$$

$$-w^\nu_{\ell,i} - \tan\left(\frac{\pi}{2^{\nu+1}}\right) \lambda^5_{\ell,i} + \lambda^6_{\ell,i} = 0,\ \forall\, \ell \in [\theta], \tag{B.5g}$$

$$-k^\nu_{\ell,i} - u^\nu_{\ell,i} + \lambda^5_{\ell,i} = 0,\ \forall\, \ell \in [\theta], \tag{B.5h}$$

$$\lambda^1_{\ell,i}, \lambda^2_{\ell,i}, \lambda^3_{\ell,i}, \lambda^4_{\ell,i}, \lambda^5_{\ell,i}, \lambda^6_{\ell,i} \geq 0,\ \forall\, \ell \in [\theta], \tag{B.5i}$$

$$k^j_{\ell,i}, u^j_{\ell,i} \geq 0,\ \forall \ell \in [\theta], j \in [\nu]. \tag{B.5j}$$

First, we consider the set

$$\mathcal{A}(M_{\theta+1}) := \left\{ \substack{\lambda^1_{\theta+1,1}, \lambda^2_{\theta+1,1}, \lambda^3_{\theta+1,1}, \lambda^4_{\theta+1,1}, \lambda^5_{\theta+1,1}, \lambda^6_{\theta+1,1}, \\ k^j_{\theta+1,1}, u^j_{\theta+1,1}, w^j_{\theta+1,1}, \forall j \in [\nu]} \,\middle|\, (\text{B.3d}) - (\text{B.3n}) \right\}.$$

By Lemma 4, for any $M_{\theta+1} > 0$, we have that $\mathcal{A}(M_{\theta+1})$ is a bounded set and $\lambda^1_{\theta+1,1}, \lambda^2_{\theta+1,1}$, $\lambda^3_{\theta+1,1}, \lambda^4_{\theta+1,1}$ are bounded variables. By (B.3d) and (B.3e), we can add two redundant constraints $\lambda^6_{\theta,1} \leq M'_\theta = 2M_{\theta+1} + 1$ and $\lambda^6_{\theta,2} \leq M'_\theta = 2M_{\theta+1} + 1$ and do not change the feasible region.

Second, we consider two sets

$$\mathcal{B}(M'_\theta) := \left\{ \substack{\lambda^1_{\ell,i}, \lambda^2_{\ell,i}, \lambda^3_{\ell,i}, \lambda^4_{\ell,i}, \lambda^5_{\ell,i}, \lambda^6_{\ell,i}, \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], \\ k^j_{\ell,i}, u^j_{\ell,i}, w^j_{\ell,i}, \forall\, \ell \in [\theta], i \in [2^{\theta-\ell}], j \in [\nu]} \;\middle|\; \text{(B.3b)}, \lambda^6_{\theta,1} \leq M'_\theta \right\},$$
$$\mathcal{C}(M'_\theta) := \left\{ \substack{\lambda^1_{\ell,i}, \lambda^2_{\ell,i}, \lambda^3_{\ell,i}, \lambda^4_{\ell,i}, \lambda^5_{\ell,i}, \lambda^6_{\ell,i}, \forall\, \ell \in [\theta], i \in \{2^{\theta-\ell}+1, \ldots, 2^{\theta+1-\ell}\}, \\ k^j_{\ell,i}, u^j_{\ell,i}, w^j_{\ell,i}, \forall\, \ell \in [\theta], i \in \{2^{\theta-\ell}+1, \ldots, 2^{\theta+1-\ell}\}, j \in [\nu]} \;\middle|\; \text{(B.3c)}, \lambda^6_{\theta,2} \leq M'_\theta \right\}.$$

Since the feasible region of Problem (B.2) is bounded for any $M_\theta > 0$, we have that $\mathcal{B}(M'_\theta)$ and $\mathcal{C}(M'_\theta)$ are bounded sets.

Finally, by the boundedness of $\mathcal{A}(M_{\theta+1})$, $\mathcal{B}(M'_\theta)$, and $\mathcal{C}(M'_\theta)$, we have that the feasible region of Problem (B.3) is bounded for any $M_{\theta+1} > 0$.

B.5 Numerical Results on Unitbox Instances

We also evaluate our MILP approximations on the unitbox instances introduced by Bao et al. (2009). Specifically, we consider the following problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^\top \mathbf{Q}_0 \mathbf{x} + \mathbf{c}_0^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^\top \mathbf{Q}_i \mathbf{x} + \mathbf{c}_i^\top \mathbf{x} \leq b_i, \ \forall\, i \in [m], \\ & 0 \leq x_i \leq 1, \ \forall\, i \in [N]. \end{aligned}$$

The number of variables N takes values from $\{50, 100, 150, 200\}$, while the number of constraints m takes values from $\{1, 2, 3\}$. All the matrices $(\mathbf{Q}_i, \forall i \in \{0, 1, \ldots, m\})$ are generated with the same density d, where d takes values from $\{1.0, 0.5, 0.25\}$. When considering $d = 1.0$, all the entries in the matrix are randomly generated. When considering $d = 0.5$, we first generate a matrix $\mathbf{Q}'_i$ with density 1.0 and then let $\mathbf{Q}_i = \begin{bmatrix} \mathbf{0} & \mathbf{Q}'_i \\ (\mathbf{Q}'_i)^\top & \mathbf{0} \end{bmatrix}$ for each $i \in \{0, 1, \ldots, m\}$. When considering $d = 0.25$, we first generate the matrix $\mathbf{Q}'_i$ by randomly selecting half of the entries in each row to be zero and then let $\mathbf{Q}_i = \begin{bmatrix} \mathbf{0} & \mathbf{Q}_i' \\ (\mathbf{Q}_i')^\top & \mathbf{0} \end{bmatrix}$ for each $i \in \{0, 1, \ldots, m\}$. All nonzero coefficients in $\mathbf{c}_i$ and $\mathbf{Q}_i$ are randomly generated on $[-1, 1]$ and the coefficient b_i is randomly chosen as integers from $[1, 100]$ for each $i \in \{0, 1, \ldots, m\}$.

To consider an appropriate number of negative eigenvalues in $\mathbf{Q}_i$ for each $i \in \{0, 1, \ldots, m\}$, we further update all the diagonal elements in $\mathbf{Q}_i$ to a positive number such that the number of negative eigenvalues in $\mathbf{Q}_i$ is less than 4. Specifically, for $N = 50$, we update all diagonal elements of all the matrices ($\mathbf{Q}_i, \forall i \in \{0, 1, \ldots, m\}$) to 4.8, 4.8, and 3.5 when $d =$ 1.0, 0.5, and 0.25, respectively. Similarly, for $N = 100$, all diagonal elements are updated to 7.5, 7.3, and 5.2 when $d = 1.0$, 0.5, and 0.25, respectively. For $N = 150$, all diagonal elements are updated to 9.2, 9.2, and 6.6 when $d = 1.0$, 0.5, and 0.25, respectively. For $N = 200$, all diagonal elements are updated to 10.7, 10.7, and 7.5 when $d = 1.0$, 0.5, and 0.25, respectively. For example, when $N = 200$ and $d = 0.25$, all diagonal elements of all the matrices ($\mathbf{Q}_i, \forall i \in \{0, 1, \ldots, m\}$) are updated to 7.5. For each combination of m, N, and d, we generate three instances and report the average performance in Tables B.2, B.3, and B.4. We set the time limit to three hours and ε to 0.01 for all the experiments. Other numerical settings and table header descriptions are provided in Section 3.4.2.

Table B.2: Average Performance of Unitbox Instances with $m = 1$

		Gurobi		LP Relaxation		SOCP Relaxation		SDP Relaxation		A1		A2	
N	d	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
50	0.25	0%/10.00%	3361/2684	58.61%	1	46.46%	1	30.94%	143	6.12%	256	4.32%	159
	0.5	0%/10.00%	933/692	56.65%	1	46.19%	1	23.20%	167	9.20%	151	5.83%	174
	1.0	0%/10.00%	5463/4082	61.45%	1	40.28%	1	33.36%	119	8.18%	287	5.28%	166
100	0.25	23.63%	10800	98.35%	1	51.03%	1	33.33%	213	8.92%	1346	4.05%	1164
	0.5	26.07%	10800	92.38%	1	44.81%	1	24.29%	220	6.88%	1356	3.83%	892
	1.0	14.24%	10800	84.94%	1	38.14%	1	31.12%	207	7.85%	445	4.45%	745
150	0.25	36.29%	10800	126.52%	1	52.78%	1	28.15%	576	7.84%	2314	5.26%	2271
	0.5	116.85%	10800	252.52%	1	117.10%	2	45.53%	519	9.00%	718	4.07%	2364
	1.0	89.58%	10800	202.66%	1	84.56%	1	56.54%	498	5.37%	3143	3.92%	2791
200	0.25	321.08%	10800	407.92%	1	167.43%	1	59.01%	1423	7.37%	9010	6.08%	8511
	0.5	271.18%	10800	441.63%	1	211.86%	1	65.46%	1627	8.94%	5482	4.62%	9421
	1.0	288.43%	10800	431.91%	1	192.72%	1	66.83%	1405	7.89%	3911	5.73%	9836

Tables B.2–B.4 show the following results. First, when $N = 50$, where the Gurobi

Table B.3: Average Performance of Unitbox Instances with $m = 2$

N	d	Gurobi		LP Relaxation		SOCP Relaxation		SDP Relaxation		A1		A2	
		Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
50	0.25	0%/10.00%	3462/2841	72.19%	1	39.62%	1	25.59%	273	7.82%	321	7.42%	192
	0.5	0%/10.00%	4355/3122	68.66%	1	42.07%	1	34.91%	197	8.91%	349	4.94%	207
	1.0	0%/10.00%	5211/4329	59.05%	1	46.17%	1	31.43%	212	7.42%	426	6.32%	385
100	0.25	25.96%	10800	113.42%	1	56.18%	1	34.71%	298	6.81%	1532	4.64%	1328
	0.5	19.15%	10800	99.83%	1	48.08%	1	31.22%	312	7.24%	1543	5.12%	815
	1.0	27.61%	10800	95.48%	1	41.11%	1	29.74%	247	7.85%	1782	3.97%	974
150	0.25	174.66%	10800	263.84%	1	87.53%	1	34.56%	623	9.21%	4325	4.82%	4893
	0.5	161.79%	10800	202.37%	1	94.21%	1	57.23%	702	8.02%	5321	5.81%	4723
	1.0	189.45%	10800	192.74%	1	69.19%	1	46.12%	539	6.61%	5322	5.94%	5481
200	0.25	266.12%	10800	412.85%	1	263.43%	1	71.83%	1849	9.24%	8234	5.61%	9712
	0.5	429.82%	10800	496.36%	1	207.11%	1	62.81%	2052	6.98%	7342	4.61%	8281
	1.0	381.38%	10800	457.07%	1	242.79%	1	83.94%	1792	7.57%	6234	7.02%	8792

Table B.4: Average Performance of Unitbox Instances with $m = 3$

N	d	Gurobi		LP Relaxation		SOCP Relaxation		SDP Relaxation		A1		A2	
		Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
50	0.25	0%/10.00%	4271/3012	69.71%	1	49.28%	1	41.29%	308	8.65%	532	5.91%	293
	0.5	0%/10.00%	4521/3472	72.94%	1	53.74%	1	38.49%	247	8.93%	491	7.22%	311
	1.0	0%/10.00%	6183/4717	60.05%	1	49.29%	1	32.08%	197	6.19%	482	4.08%	297
100	0.25	27.41%	10800	152.71%	1	84.28%	1	40.11%	402	7.89%	2014	4.92%	1693
	0.5	31.29%	10800	87.62%	1	51.29%	1	33.85%	384	6.19%	1974	6.03%	2503
	1.0	25.41%	10800	95.82%	1	61.72%	2	30.59%	292	8.42%	2285	2.93%	1649
150	0.25	192.41%	10800	252.60%	1	102.71%	1	54.82%	704	6.94%	5832	4.92%	6831
	0.5	201.52%	10800	249.61%	1	88.91%	1	54.48%	692	8.82%	5732	6.16%	8319
	1.0	228.64%	10800	309.82%	1	72.71%	1	40.43%	815	7.18%	6294	4.05%	7406
200	0.25	494.45%	10800	552.84%	2	225.04%	1	88.02%	2081	9.01%	7295	7.21%	10039
	0.5	374.81%	10800	482.44%	2	188.07%	1	59.15%	2207	9.12%	9451	6.62%	10518
	1.0	422.38%	10800	501.02%	1	215.74%	1	73.59%	2498	7.82%	9703	4.53%	9984

solver can solve each instance to optimality (see the column "Gurobi"), our proposed approximations exhibit efficient performance in obtaining lower bounds (see the columns "A1" and "A2"). On the one hand, our proposed two approximations require less time to provide better lower bounds (with smaller gaps) than the Gurobi solver. For example, when $m = 1$, $N = 50$, and $d = 0.25$, the Gurobi solver takes 2684 seconds to obtain a solution with a gap of 10.00%, but our second approximation A2 only takes 159 seconds to obtain a solution with a gap of 4.32%. On the other hand, our proposed two approximations show significant improvement in the optimality gap compared to the LP, SOCP, and SDP relaxations of the original non-convex QCP (3.1), where the relaxations need less time to be solved. Moreover, A2 achieves smaller gaps quicker than our first approximation A1.

Second, when $N \geq 100$, where the Gurobi solver cannot solve all the instances within the time limit of three hours, our proposed approximations still deliver tight lower bounds. In such cases, our approximations ensure that the gap remains below 10.00%. For instance, when $m = 1$, $N = 200$, and $d = 1.0$, the Gurobi solver obtains a solution with a gap of 288.43% within three hours, but A1 can obtain a solution with a gap of 7.89% in 3911 seconds. Moreover, when $N = 200$, A2 requires more computational time to solve than A1.

Third, A2 generally outperforms A1 in terms of the optimality gap. However, when N increases, the computational time of A2 grows faster than A1. These findings align with the conclusions drawn from our randomly generated instances in Section 3.4.2 and also reflect our discussions at the end of Section 3.3 that both approximations, with their respective advantages and disadvantages, can be useful and relevant.

Appendix C

Supplements for Chapter 4

C.1 Table of Notation

Table C.1: Summary of Notations

Notation	Description
Indices and Sets:	
$n/k/r$	Index of buses/ renewable distributed generation (RDG) units/ energy storage (ES) units.
$\pi/\kappa/\tau$	Index of different data samples.
$\mathcal{N}/\mathcal{E}/\mathcal{N}_n$	Set of all the buses/ all the power distribution lines/ the buses connected to a given Bus $n \in \mathcal{N}$, i.e., $\mathcal{N}_n := \{m \mid (n,m) \in \mathcal{E}\}$.
$\mathcal{B}_0/\mathcal{B}_1/\mathcal{B}_2$	Set of buses that are connected to Bus 0/ installed with dispatchable distributed generation (DDG) units/ installed with reactive power sources.
$\mathcal{X}$	Set of standard capacities of RDG units, i.e., $\mathcal{X} := \{\bar{x}_1, \ldots, \bar{x}_L\}$.
K	$\{1, 2, \ldots, K\}$, for any $K \in \mathbb{Z}_+$.
Parameters:	

c^0_{kn}/d^0_{rn}	Setup cost of placing the k^{th} candidate RDG unit/ the r^{th} candidate ES unit at Bus n.
c^1_k/c^2_k	Size-based investment/ maintenance cost of the k^{th} candidate RDG unit.
d^1_r/d^2_r	Size-based investment/ maintenance cost of the r^{th} candidate ES unit.
c^t_p/c^t_q	Electricity price of purchasing active/reactive power from the main grid via Bus 0.
c^f_n/c^e_n	Fuel/emission price for the DDG units at Bus n.
ω	Emission factor of the DDG units (kg/kWh).
$K/R/T/N$	Total number of candidate RDG units to be installed/ candidate ES units to be installed/ all the time intervals in the planning horizon/ all the buses.
$\bar{K}$	Maximum number of RDG units to be installed.
LC_{mn}	Capacity of a power distribution line $(m,n) \in \mathcal{E}$.
$(\bar{p}^t_n, \underline{p}^t_n)$	Active power output bounds of DDG unit n in period t.
$(\bar{q}^t_n, \underline{q}^t_n)$	Reactive power output bounds of reactive power source n in period t.
$\mathfrak{R}_{mn}/\mathfrak{X}_{mn}$	Electrical resistance/reactance of line (m,n).
δ_n/τ_n	Binary indicator of whether a DDG unit/ a reactive power source is at Bus n.
$\underline{v}/\bar{v}$	Upper/lower bound of voltage magnitude at a bus.
$\bar{y}_r/\underline{y}_r$	Maximum/minimum capacity of the r^{th} candidate ES unit.
e_1/e_2	ES charging/discharging unit cost.
$\eta/\gamma/b^0_r$	Violation probability/ ES efficiency/ Initial power level of the r^{th} candidate ES unit.
$\Pi_1/\Pi_2/\Pi_3$	Total number of data samples of different types.

Random Variables:

d_{pn}^t/d_{qn}^t	Active/reactive load at Bus n in period t.
s_k^t	Active power output efficiency of the k^{th} candidate RDG unit in period t.
$\boldsymbol{\xi}^t$	Vector of uncertainty in compact form in period t, i.e., $[d_{p1}^t, \ldots, d_{pN}^t, d_{q1}^t, \ldots, d_{qN}^t, s_1^t, \ldots, s_k^t]^\top$.
Decision Variables:	
$z_{kn}/u_{kl}/w_{rn}$	Binary indicator of whether the k^{th} candidate RDG unit is located at Bus n/ of whether the capacity of the k^{th} candidate RDG unit is the l^{th} element in $\mathcal{X}$/ of whether the r^{th} candidate ES unit is located at Bus n.
x_k/y_r	Size of the k^{th} candidate RDG unit/ Capacity of the r^{th} candidate ES unit.
f_r^t/g_r^t	Active power that is charged/ discharged at the r^{th} candidate ES unit in period t.
p_0^t/q_0^t	Active/ reactive power purchased from the main grid via Bus 0 in period t.
P_{mn}^t/Q_{mn}^t	Active/ reactive power flow from Bus m to n in period t.
$V_n^t/\|V_n^t\|^2$	Complex voltage at Bus n in period t/ its magnitude.
$I_{mn}^t/\|I_{mn}^t\|^2$	Complex current from Bus m to n in period t/ its magnitude.
p_n^t/q_n^t	Active/ reactive power output of the DDG unit/ reactive power source at Bus n in period t.
b_r^t	Active power of the r^{th} candidate ES unit in period t.
LS_{1mn}^t/LS_{2mn}^t	Load-shedding variables.
$\boldsymbol{z}$	$[z_{kn}, \forall\, k \in [K], n \in \mathcal{N}]^\top$.
$\boldsymbol{x},\ \boldsymbol{u}$	$[x_1, \ldots, x_K]^\top$, $[u_{kl}, \forall\, k \in [K], l \in [L]]^\top$.
$\boldsymbol{y}, \boldsymbol{w}$	$[y_1, \ldots, y_R]^\top$, $[w_{rn}, \forall\, r \in [R], n \in \mathcal{N}]^\top$.

References

Abbey, C. and Joos, G. (2007). Supercapacitor energy storage for wind energy applications. *IEEE Transactions on Industry Applications*, 43(3):769–776.

Ackermann, T., Andersson, G., and Söder, L. (2001). Distributed generation: a definition. *Electric Power Systems Research*, 57(3):195–204.

Al-Khayyal, F. A., Larsen, C., and Van Voorhis, T. (1995). A relaxation method for nonconvex quadratically constrained quadratic programs. *Journal of Global Optimization*, 6(3):215–230.

Audet, C., Hansen, P., Jaumard, B., and Savard, G. (2000). A branch and cut algorithm for nonconvex quadratically constrained quadratic programming. *Mathematical Programming*, 87:131–152.

Bach, F. R., Lanckriet, G. R., and Jordan, M. I. (2004). Multiple kernel learning, conic duality, and the smo algorithm. In *Proceedings of the 21st International Conference on Machine Learning*, page 6.

Bai, X., Sun, J., and Zheng, X. (2021). An augmented lagrangian decomposition method for chance-constrained optimization problems. *INFORMS Journal on Computing*, 33(3):1056–1069.

Ban, G.-Y., El Karoui, N., and Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, 64(3):1136–1154.

Ban, G.-Y. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108.

Bao, X., Sahinidis, N. V., and Tawarmalani, M. (2009). Multiterm polyhedral relaxations for nonconvex, quadratically constrained quadratic programs. *Optimization Methods and Software*, 24(4-5):485–504.

Bao, X., Sahinidis, N. V., and Tawarmalani, M. (2011). Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Mathematical Programming*, 129:129–157.

Baran, M. and Wu, F. F. (1989). Optimal sizing of capacitors placed on a radial distribution system. *IEEE Transactions on Power Delivery*, 4(1):735–743.

Barton, J. and Infield, D. (2004). Energy storage and its use with intermittent renewable energy. *IEEE Transactions on Energy Conversion*, 19(2):441–448.

Bazrafshan, M. and Gatsis, N. (2017). Comprehensive modeling of three-phase distribution systems via the bus admittance matrix. *IEEE Transactions on Power Systems*, 33(2):2015–2029.

Belotti, P., Lee, J., Liberti, L., Margot, F., and Wächter, A. (2009). Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24(4-5):597–634.

Ben-Tal, A. and Nemirovski, A. (2001). On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.

Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.

Bertsimas, D., Doan, X. V., Natarajan, K., and Teo, C.-P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602.

Bertsimas, D., Jaillet, P., and Martin, S. (2019a). Online vehicle routing: The edge of optimization in large-scale applications. *Operations Research*, 67(1):143–162.

Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

Bertsimas, D. and Koduri, N. (2022). Data-driven optimization: A reproducing kernel Hilbert space approach. *Operations Research*, 70(1):454–471.

Bertsimas, D., Sim, M., and Zhang, M. (2019b). Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618.

Bestuzheva, K., Besançon, M., Chen, W.-K., Chmiela, A., Donkiewicz, T., van Doornmalen, J., Eifler, L., Gaul, O., Gamrath, G., Gleixner, A., et al. (2021). The SCIP optimization suite 8.0. *arXiv preprint arXiv:2112.08872*.

Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science and Business Media.

Burer, S. (2009). On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120(2):479–495.

Burer, S. and Anstreicher, K. M. (2013). Second-order-cone constraints for extended trust-region subproblems. *SIAM Journal on Optimization*, 23(1):432–451.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357.

Burer, S. and Vandenbussche, D. (2008). A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations. *Mathematical Programming*, 113(2):259–282.

Burer, S. and Yang, B. (2015). The trust region subproblem with non-intersecting linear constraints. *Mathematical Programming*, 149(1-2):253–264.

Cain, M. B., O'neill, R. P., Castillo, A., et al. (2012). History of optimal power flow and formulations. *Federal Energy Regulatory Commission*, 1:1–36.

Celli, G., Ghiani, E., Mocci, S., and Pilo, F. (2005). A multiobjective evolutionary algorithm for the sizing and siting of distributed generation. *IEEE Transactions on Power Systems*, 20(2):750–757.

Chen, L., Fu, C., Si, F., Sim, M., and Xiong, P. (2023). Robust optimization with moment-dispersion ambiguity. *Available at SSRN 4525224*.

Chen, Y. and Goulart, P. (2023). Burer-Monteiro ADMM for large-scale SDPs. *arXiv preprint arXiv:2302.04016*.

Cheng, J., Delage, E., and Lisser, A. (2014). Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506.

Cheng, J., Gicquel, C., and Lisser, A. (2019). Partial sample average approximation method for chance constrained problems. *Optimization Letters*, 13(4):657–672.

Cheng, J., Li-Yang Chen, R., Najm, H. N., Pinar, A., Safta, C., and Watson, J.-P. (2018). Distributionally robust optimization with principal component analysis. *SIAM Journal on Optimization*, 28(2):1817–1841.

Cheramin, M., Cheng, J., Jiang, R., and Pan, K. (2022). Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity. *INFORMS Journal on Computing*, 34(3):1768–1794.

Daratha, N., Das, B., and Sharma, J. (2013). Coordination between OLTC and SVC for voltage regulation in unbalanced distribution system distributed generation. *IEEE Transactions on Power Systems*, 29(1):289–299.

De Maio, A., Huang, Y., Palomar, D. P., Zhang, S., and Farina, A. (2010). Fractional QCQP with applications in ML steering direction estimation for radar detection. *IEEE Transactions on Signal Processing*, 59(1):172–185.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.

Dong, H. and Luo, Y. (2018). Compact disjunctive approximations to nonconvex quadratically constrained programs. *arXiv preprint arXiv:1811.08122*.

Ehsan, A. and Yang, Q. (2018). Optimal integration and planning of renewable distributed generation in the power distribution networks: A review of analytical techniques. *Applied Energy*, 210:44–59.

Eldén, L. and Park, H. (1999). A procrustes problem on the stiefel manifold. *Numerische Mathematik*, 82(4):599–619.

Elmachtoub, A. N. and Grigas, P. (2022). Smart “predict, then optimize”. *Management Science*, 68(1):9–26.

Farivar, M. and Low, S. H. (2013). Branch flow model: Relaxations and convexification—part I. *IEEE Transactions on Power Systems*, 28(3):2554–2564.

Farnia, F. and Tse, D. (2016). A minimax approach to supervised learning. *Advances in Neural Information Processing Systems*, 29.

Fathabad, A. M., Cheng, J., Pan, K., and Qiu, F. (2020). Data-driven planning for renewable distributed generation integration. *IEEE Transactions on Power Systems*, 35(6):4357–4368.

Fathabad, A. M., Cheng, J., Pan, K., and Yang, B. (2023). Asymptotically tight conic approximations for chance-constrained AC optimal power flow. *European Journal of Operational Research*, 305(2):738–753.

Gallego, G. and Moon, I. (1993). The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society*, 44(8):825–834.

Gan, L. and Low, S. H. (2014). Convex relaxations and linear approximation for optimal power flow in multiphase radial networks. In *2014 Power Systems Computation Conference*, pages 1–9. IEEE.

Georgilakis, P. S. and Hatziargyriou, N. D. (2013). Optimal distributed generation placement in power distribution networks: Models, methods, and future research. *IEEE Transactions on Power Systems*, 28(3):3420–3428.

Ghaoui, L. E., Oks, M., and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556.

Ghosal, S. and Wiesemann, W. (2020). The distributionally robust chance-constrained vehicle routing problem. *Operations Research*, 68(3):716–732.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

Gotoh, J.-y. and Konno, H. (2002). Bounding option prices by semidefinite programming: A cutting plane algorithm. *Management Science*, 48(5):665–678.

Gurobi (2024). *Gurobi Optimizer Reference Manual*. Accessed: April 5, 2024.

Hajinezhad, D. and Shi, Q. (2018). Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications. *Journal of Global Optimization*, 70(1):261–288.

He, L. and Mak, H.-Y. (2023). Prescriptive PCA: Dimensionality reduction for two-stage stochastic optimization. *arXiv preprint arXiv:2306.02223*, Jun 4, 2023.

Helmberg, C. (2002). Semidefinite programming. *European Journal of Operational Research*, 137(3):461–482.

Ho-Nguyen, N. and Kılınç-Karzan, F. (2022). Risk guarantees for end-to-end prediction and optimization processes. *Management Science*, 68(12):8680–8698.

Huang, Y. and Palomar, D. P. (2014). Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing. *IEEE Transactions on Signal Processing*, 62(5):1093–1108.

Jiang, R. and Li, D. (2019). Second order cone constrained convex relaxations for nonconvex quadratically constrained quadratic programming. *Journal of Global Optimization*, 75(2):461–494.

Jiang, S., Cheng, J., Pan, K., and Shen, Z.-J. M. (2023). Optimized dimensionality reduction for moment-based distributionally robust optimization. *arXiv preprint arXiv:2305.03996*, May 6, 2023.

Jiang, S., Cheng, J., Pan, K., and Shen, Z.-J. M. (2025a). Optimized dimensionality reduction for moment-based distributionally robust optimization. Available for download at https://github.com/jsy1164014200/ODR-MDRO.

Jiang, S., Cheng, J., Pan, K., and Yang, B. (2025b). Asymptotically tight MILP approximations for a non-convex QCP. Available for download at https://doi.org/10.1287/ijoc/2024.0719.cd, https://github.com/INFORMSJoC/2024.0719.

Kallus, N. and Mao, X. (2023). Stochastic optimization forests. *Management Science*, 69(4):1935–2545.

Kannan, R., Nagarajan, H., and Deka, D. (2022). Learning to accelerate the global optimization of quadratically-constrained quadratic programs. *arXiv preprint arXiv:2301.00306*.

Karimi, R., Cheng, J., and Lejeune, M. A. (2021). A framework for solving chance-constrained linear matrix inequality programs. *INFORMS Journal on Computing*, 33(3):1015–1036.

Keane, A., Ochoa, L. F., Borges, C. L. T., Ault, G. W., Alarcon-Rodriguez, A. D., Currie, R. A. F., Pilo, F., Dent, C., and Harrison, G. P. (2013). State-of-the-art techniques and challenges ahead for distributed generation planning and optimization. *IEEE Transactions on Power Systems*, 28(2):1493–1502.

Kim, S. and Kojima, M. (2001). Second order cone programming relaxation of nonconvex quadratic optimization problems. *Optimization Methods and Software*, 15(3-4):201–224.

Kuang, X., Ghaddar, B., Naoum-Sawaya, J., and Zuluaga, L. F. (2016). Alternative LP and SOCP hierarchies for ACOPF problems. *IEEE Transactions on Power Systems*, 32(4):2828–2836.

Lanckriet, G. R., Ghaoui, L. E., Bhattacharyya, C., and Jordan, M. I. (2002). A robust minimax approach to classification. *Journal of Machine Learning Research*, 3(Dec):555–582.

Lemon, A., So, A. M.-C., Ye, Y., et al. (2016). Low-rank semidefinite programming: Theory and applications. *Foundations and Trends® in Optimization*, 2(1-2):1–156.

Li, J. Y.-M. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, 66(6):1533–1541.

Lin, F., Fang, X., and Gao, Z. (2022). Distributionally robust optimization: A review on theory and applications. *Numerical Algebra, Control and Optimization*, 12(1):159–212.

Lin, Y. and Schrage, L. (2009). The global solver in the LINDO API. *Optimization Methods and Software*, 24(4-5):657–668.

Linderoth, J. (2005). A simplicial branch-and-bound algorithm for solving quadratically constrained quadratic programs. *Mathematical Programming*, 103:251–282.

Liu, S., He, L., and Shen, Z.-J. M. (2021). On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, 67(7):4095–4119.

Liu, X., Küçükyavuz, S., and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, 157(1):219–243.

Liu, Y., Meskarian, R., and Xu, H. (2017). Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM Journal on Optimization*, 27(2):957–985.

Löfberg, J. (2004). Yalmip: A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan.

Lotfi, S. and Zenios, S. A. (2018). Robust VaR and CVaR optimization under joint ambiguity in distributions, means, and covariances. *European Journal of Operational Research*, 269(2):556–576.

Low, S. H. (2014). Convex relaxation of optimal power flow—part I: Formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27.

Lubin, M., Dvorkin, Y., and Roald, L. (2019). Chance constraints for improving the security of AC optimal power flow. *IEEE Transactions on Power Systems*, 34(3):1908–1917.

Luo, H., Sun, X., and Li, D. (2008). On the convergence of augmented lagrangian methods for constrained global optimization. *SIAM Journal on Optimization*, 18(4):1209–1230.

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I—convex underestimating problems. *Mathematical Programming*, 10(1):147–175.

Mehrotra, S. and Papp, D. (2014). A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, 24(4):1670–1697.

Melgar-Dominguez, O. D., Pourakbari-Kasmaei, M., and Mantovani, J. R. S. (2018). Adaptive robust short-term planning of electrical distribution systems considering siting and sizing of renewable energy based dg units. *IEEE Transactions on Sustainable Energy*, 10(1):158–169.

Misener, R. and Floudas, C. A. (2012). Global optimization of mixed-integer quadratically-constrained quadratic programs (MIQCQP) through piecewise-linear and edge-concave relaxations. *Mathematical Programming*, 136(1):155–182.

Misener, R. and Floudas, C. A. (2013). Glomiqo: Global mixed-integer quadratic optimizer. *Journal of Global Optimization*, 57(1):3–50.

Misener, R. and Floudas, C. A. (2014). Antigone: algorithms for continuous/integer global optimization of nonlinear equations. *Journal of Global Optimization*, 59(2-3):503–526.

Mohajerin Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166.

Natarajan, K., Sim, M., and Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science*, 64(7):3146–3167.

Nemirovski, A. and Shapiro, A. (2007). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996.

Oree, V., Hassen, S. Z. S., and Fleming, P. J. (2017). Generation expansion planning optimisation with renewable energy integration: A review. *Renewable and Sustainable Energy Reviews*, 69:790–803.

Ozturk, U. A., Mazumdar, M., and Norman, B. A. (2004). A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Transactions on Power Systems*, 19(3):1589–1598.

Pagnoncelli, B. K., Ahmed, S., and Shapiro, A. (2009). Sample average approximation method for chance constrained programming: theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416.

Pandžić, H., Wang, Y., Qiu, T., Dvorkin, Y., and Kirschen, D. (2014). Near-optimal method for siting and sizing of distributed storage in a transmission network. *IEEE Transactions on Power Systems*, 30(5):2288–2300.

Park, H., Baldick, R., and Morton, D. P. (2015). A stochastic transmission planning model with dependent load and wind forecasts. *IEEE Transactions on Power Systems*, 30(6):3003–3011.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Poljak, S., Rendl, F., and Wolkowicz, H. (1995). A recipe for semidefinite relaxation for (0, 1)-quadratic programming: In memory of Svata Poljak. *Journal of Global Optimization*, 7(1):51–73.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112.

Pozo, D. and Contreras, J. (2012). A chance-constrained unit commitment with an $n - k$ security criterion and significant wind generation. *IEEE Transactions on Power Systems*, 28(3):2842–2851.

Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., and Shen, Z.-J. M. (2023). A practical end-to-end inventory management model with deep learning. *Management Science*, 69(2):759–773.

Qian, K., Zhou, C., Allan, M., and Yuan, Y. (2010). Modeling of load demand due to EV battery charging in distribution systems. *IEEE Transactions on Power Systems*, 26(2):802–810.

Qin, Z., Li, W., and Xiong, X. (2012). Generation system reliability evaluation incorporating correlations of wind speeds with different distributions. *IEEE Transactions on Power Systems*, 28(1):551–558.

Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, Aug 13, 2019.

Rahimian, H. and Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85.

Roald, L. and Andersson, G. (2017). Chance-constrained AC optimal power flow: Reformulations and efficient algorithms. *IEEE Transactions on Power Systems*, 33(3):2906–2918.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. (2016). Robust growth-optimal portfolios. *Management Science*, 62(7):2090–2109.

Sahinidis, N. V. (1996). Baron: A general purpose global optimization software package. *Journal of Global Optimization*, 8:201–205.

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production.*

Shehadeh, K. S., Cohn, A. E., and Jiang, R. (2020). A distributionally robust optimization approach for outpatient colonoscopy scheduling. *European Journal of Operational Research*, 283(2):549–561.

Shen, Y., Wen, Z., and Zhang, Y. (2014). Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optimization Methods and Software*, 29(2):239–263.

Sherali, H. D. (2007). RLT: A unified approach for discrete and continuous nonconvex optimization. *Annals of Operations Research*, 149(1):185.

Sherali, H. D. and Adams, W. P. (2013). *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, volume 31. Springer Science & Business Media.

Shor, N. Z. (1987). Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11.

Shor, N. Z. (1990). Dual quadratic estimates in polynomial and boolean programming. *Annals of Operations Research*, 25(1):163–168.

Shor, N. Z. (1992). Dual estimates in multiextremal problems. *Journal of Global Optimization*, 2:411–418.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176.

So, A. M.-C. (2011). Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical programming*, 130(1):125–151.

Sturm, J. F. and Zhang, S. (2003). On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28(2):246–267.

Tan, Y., Cao, Y., Li, C., Li, Y., Zhou, J., and Song, Y. (2014). A two-stage stochastic programming approach considering risk level for distribution networks operation with wind power. *IEEE Systems Journal*, 10(1):117–126.

Tawarmalani, M. and Sahinidis, N. V. (2004). Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Mathematical Programming*, 99(3):563–591.

Themelis, A. and Patrinos, P. (2020). Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Review*, 38(1):49–95.

Vielma, J. P., Dunning, I., Huchette, J., and Lubin, M. (2017). Extended formulations in mixed integer conic quadratic programming. *Mathematical Programming Computation*, 9:369–418.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wang, Q., Guan, Y., and Wang, J. (2012). A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. *IEEE Transactions on Power Systems*, 27(1):206–215.

Wang, Y., Yin, W., and Zeng, J. (2019). Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78:29–63.

Wolkowicz, H. (2000). Semidefinite and Lagrangian relaxations for hard combinatorial problems. In *System Modelling and Optimization: Methods, Theory and Applications. 19 th IFIP TC7 Conference on System Modelling and Optimization July 12–16, 1999, Cambridge, UK 19*, pages 269–309. Springer.

Wu, H., Shahidehpour, M., Li, Z., and Tian, W. (2014). Chance-constrained day-ahead scheduling in stochastic power system operation. *IEEE Transactions on Power Systems*, 29(4):1583–1591.

Wu, Z., Zeng, P., Zhang, X.-P., and Zhou, Q. (2016). A solution to the chance-constrained two-stage stochastic program for unit commitment with wind integration. *IEEE Transactions on Power Systems*, 31(6):4185–4196.

Xavier, Á. S., Qiu, F., and Ahmed, S. (2021). Learning to solve large-scale security-constrained unit commitment problems. *INFORMS Journal on Computing*, 33(2):739–756.

Xia, W., Vera, J. C., and Zuluaga, L. F. (2020). Globally solving nonconvex quadratic programs via linear integer programming techniques. *INFORMS Journal on Computing*, 32(1):40–56.

Xiang, Z. and Tao, M. (2012). Robust beamforming for wireless information and power transmission. *IEEE Wireless Communications Letters*, 1(4):372–375.

Xu, H., Liu, Y., and Sun, H. (2018). Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming*, 169(2):489–529.

Yue, J., Chen, B., and Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136.

Yurtsever, A., Tropp, J. A., Fercoq, O., Udell, M., and Cevher, V. (2021). Scalable semidefinite programming. *SIAM Journal on Mathematics of Data Science*, 3(1):171–200.

Zang, Y. and Zhu, H. (2022a). An eigenvalue decomposition method for low-rank and non-convex quadratically constrained quadratic programming. In *2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 558–563. IEEE.

Zang, Y. and Zhu, H. (2022b). Fast and optimal joint decision and estimation by quantized data via noisy channels of sensor networks. *Signal Processing*, 195:108481.

Zeng, B., An, Y., and Kuznia, L. (2014). Chance constrained mixed integer program: Bilinear and linear formulations, and benders decomposition. *arXiv preprint arXiv:1403.7875*.

Zhang, L., Yang, J., and Gao, R. (2023). Optimal robust policy for feature-based newsvendor. *Management Science*, Forthcoming.

Zhang, S., Cheng, H., Zhang, L., Bazargan, M., and Yao, L. (2013). Probabilistic evaluation of available load supply capability for distribution system. *IEEE Transactions on Power Systems*, 28(3):3215–3225.

Zhang, Y., Song, S., Shen, Z.-J. M., and Wu, C. (2017a). Robust shortest path problem with distributional uncertainty. *IEEE Transactions on Intelligent Transportation Systems*, 19(4):1080–1090.

Zhang, Y., Wang, J., Zeng, B., and Hu, Z. (2017b). Chance-constrained two-stage unit commitment under uncertain load and wind output using bilinear benders. *IEEE Transactions on Power Systems*, 32(5):3637–3647.

Zhang, Y. J. A. and So, A. M.-C. (2011). Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming. *IEEE Journal on Selected Areas in Communications*, 29(2):362–373.

Zhou, Z., Zhang, J., Liu, P., Li, Z., Georgiadis, M., and Pistikopoulos, E. (2013). A two-stage stochastic programming model for the optimal design of distributed energy systems. *Applied Energy*, 103:135–144.

Zhu, T., Xie, J., and Sim, M. (2022). Joint estimation and robustness optimization. *Management Science*, 68(3):1659–1677.

Zien, A. and Ong, C. S. (2007). Multiclass multiple kernel learning. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1191–1198.

Zymler, S., Kuhn, D., and Rustem, B. (2013). Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188.